

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OW nucleic - protein search, using frame_plus_n2p model

Run on: January 29, 2004, 23:39:31 ; Search time 125 Seconds

(without alignments)
12886.115 Million cell updates/sec

Title: US-10-032-996-6
Perfect score: 5739
Sequence: 1 ggcgccttgagctccgcctcc.....aaaaaatcttgaagagac 3121

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+np.model -DEV=xi1p
-O=/cgn2.1/USPTO_spo1_p/US10033996/runatc_29012004_160249_24126/app_query.fasta_1.3271
-DB=SPREMBL_23 -QPM=fastran -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCU=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10033996 @CGN 1.1 168 @runatc_29012004_160249_24126 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT_DSPBLOCK=100 -LONGIOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THRAIDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SPREMBL_23:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriophage:*
17: SP_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2334	40.7	468	4 Q81W72	Q81W72 homo sapien

2	1250.5	21.8	833	4 Q9UF54	Q9UF54 homo sapien
3	1243.5	21.7	669	11 Q8WB89	Q8WB89 mus musculus
4	626.5	11.0	676	6 Q95UC9	Q95UC9 sus scrofa
5	598	10.4	1953	5 Q9B1T7	Q9B1T7 nephila mad
6	560.5	9.8	1453	13 Q9Y1B4	Q9Y1B4 cynops pyr
7	560.5	9.8	1453	11 Q63079	Q63079 ratus norv
8	553	9.7	950	5 Q9VE45	Q9VE45 drosophila
9	548.5	9.6	566	6 Q9C7D1	Q9C7D1 sus scrofa
10	541.5	9.5	1099	5 Q81B84	Q81B84 drosophila
11	540.5	9.5	1461	4 Q76045	Q76045 homo sapien
12	540.5	9.5	1464	4 Q8N473	Q8N473 homo sapien
13	533	9.3	1447	13 Q91B91	Q91B91 xenopus lae
14	532.5	9.3	1464	11 Q8BLM4	Q8BLM4 mus musculus
15	531.5	9.3	1464	11 Q8BXY2	Q8BXY2 mus musculus
16	527	9.2	1447	13 Q91B91	Q91B91 xenopus lae
17	523	9.2	1445	13 Q93251	Q93251 rana catesb
18	522	9.1	1445	13 Q93251	Q93251 rana catesb
19	518.5	9.0	1449	13 Q910C0	Q910C0 oncorhynch
20	517	9.0	1487	4 Q14047	Q14047 homo sapien
21	511	8.9	1160	4 Q14046	Q14046 homo sapien
22	511	8.9	1430	13 Q90W37	Q90W37 gallus gall
23	509.5	8.9	1461	4 Q8N473	Q8N473 homo sapien
24	509.5	8.9	1464	4 Q8N473	Q8N473 homo sapien
25	509	8.9	2715	5 Q61603	Q61603 muric herpe
26	509	8.9	774	12 Q41971	Q41971 muric herpe
27	508	8.9	1442	11 Q62031	Q62031 mus musculus
28	508	8.9	1442	11 Q62033	Q62033 mus musculus
29	508	8.9	1459	11 Q62032	Q62032 mus musculus
30	508	8.9	1459	11 Q62032	Q62032 mus musculus
31	506	8.9	1160	4 Q14046	Q14046 homo sapien
32	505.5	8.8	1346	13 Q8U0J3	Q8U0J3 oncorhynch
33	504.5	8.8	1352	13 Q8U0J4	Q8U0J4 oncorhynch
34	504.5	8.8	1453	11 Q63079	Q63079 ratus norv
35	504.5	8.8	2703	5 Q9VEG7	Q9VEG7 drosophila
36	504.5	8.8	2716	5 Q81N94	Q81N94 drosophila
37	504	8.8	1458	13 Q910B9	Q910B9 oncorhynch
38	504	8.8	3084	12 Q8U211	Q8U211 pseudorabie
39	502	8.8	1449	13 Q910C0	Q910C0 oncorhynch
40	501.5	8.8	1953	5 Q9B1T7	Q9B1T7 nephila mad
41	501	8.8	1487	6 Q77753	Q77753 canis fam11
42	500.5	8.7	1418	6 Q28396	Q28396 equus cabal
43	500.5	8.7	1419	5 Q97406	Q97406 halioctis di
44	498.5	8.7	1414	5 Q26634	Q26634 strongyloce
45	498	8.7	1442	11 Q62031	Q62031 mus musculus

ALIGNMENTS

RESULT 1	ID	Q81W72	PRELIMINARY:	PRT:	468 AA.
AC	Q81W72				
DT	01-MAR-2003 (TRENBLrel. 23, Created)				
DT	01-MAR-2003 (TRENBLrel. 23, last sequence update)				
DT	01-MAR-2003 (TRENBLrel. 23, last annotation update)				
DE	Hypothetical protein (Fragment)				
OC	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Ovary;				
RA	Strasberg R;				
RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC040878; AAH40878.1;				
KW	Hypothetical protein.				
FT	NON TER				
SQ	SEQUENCE	468 AA;	54032 MW;	CE4172079C033B92 CRC64;	

Alignment Scores:
Pred. No.: 2,334-100
Score: 2334.00
Length: 468
Matches: 466

Percent Similarity: 99.57% Conservative: 0
 Best Local Similarity: 99.57% Mismatches: 2
 Query Match: 40.67% Indels: 0
 DB: 4 Gaps: 0

US-10-032-996-6 (1-3121) x Q81W72 (1-468)

QY 1272 ATCTGAATATTAAAGCCGAGAGATCCGGCTTCAGAGAGAGAGCCGATGTCAGC 1331
 Db 1 LLeuLysLLeuYshrdLugLulLeaLaalPheGlnArgLysArgArgSerGlySer 20
 QY 1332 AACGGCTCTGTGCTCAGCTTGAGAACAGCAGAGAGATTGAGAGCAGAGAGAGAG 1391
 Db 21 AsnGlySerValValSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
 QY 1392 GACGAGAGATGAGAGAGAGTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1451
 Db 41 AspGlnGlnMetGlnLysValLLeuGlnGlnArgArgLLeuGlnGlnGlnGln 60
 QY 1452 CTCACAG 1511
 Db 61 LeuHisLysArgGlnLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuL 80
 QY 1512 CTGAG 1571
 Db 81 LeuGlnLLeuLysArgLLeuLysArgSerGlnLLeuLLeuLLeuLLeuLLeu 100
 QY 1572 AGCGAG 1631
 Db 101 SerArgLeuGlnLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuL 120
 QY 1632 AGCGAG 1691
 Db 121 SerArgLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 140
 QY 1692 TCGCTGCTCAAG 1751
 Db 141 SerLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 160
 QY 1752 CCGAG 1811
 Db 161 ProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 180
 QY 1812 ATGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 1871
 Db 181 LLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 200
 QY 1872 TTGCTGCTCAAG 1931
 Db 201 LeuLeuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220
 QY 1932 ACCAG 1991
 Db 221 ThrArgLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 240
 QY 1992 CAG 2051
 Db 241 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
 QY 2052 TACTGCTGAG 2111
 Db 261 TyrTrpLeuGlnValLLeuGlnLLeuGlnLLeuGlnLLeuGlnLLeuGlnL 280
 QY 2112 CAG 2171
 Db 281 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
 QY 2172 GGTGAAG 2231
 Db 301 GlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
 QY 2232 GAAGTGGAG 2291

Db 321 GlnLeuLysArgGlyMetTrpLLeuLLeuLLeuLLeuLLeuLLeuLLeuL 340
 QY 2292 GCTGTAG 2351
 Db 341 AlaValGlyHisSerArgGlyGlyGlnLysArgSerLeuLysSerGlnGlnGln 360
 QY 2352 CCGTGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2411
 Db 361 ProLysMetLysPheLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 380
 QY 2412 GAGGAG 2471
 Db 381 GlnGlyLysLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuL 400
 QY 2472 ACCGTGAAG 2531
 Db 401 ThrTrpLysArgSerSerLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuL 420
 QY 2532 CCGAG 2591
 Db 421 ArgGlnLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 440
 QY 2592 TCGAG 2651
 Db 441 TrpAsnPheGlyProLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 460
 QY 2652 ATTGATGTCGAG 2711
 Db 461 LLeuLysValArgLysAsnProLeu 480

RESULT 2
 Q9UP54 PRELIMINARY; PRT; 833 AA.
 ID Q9UP54
 AC Q9UP54
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN DKF2P434D0917.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Bloecher H., Boecher M., Brandt P., Newes H.W., Gassenhuber J.,
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL133654; CAB63770.1; -.
 KW Hypothetical protein.
 FT NON-TER
 SQ SEQUENCE 833 AA; 96688 MW; 72D59966CFB177F3 CRC64;

Alignment Scores:
 Pred. No.: 2,79e-50 Length: 833
 Score: 1250.50 Matches: 291
 Percent Similarity: 65.23% Conservative: 148
 Best Local Similarity: 43.24% Mismatches: 197
 Query Match: 21.79% Indels: 37
 DB: 4 Gaps: 10

US-10-032-996-6 (1-3121) x Q9UP54 (1-833)

QY 741 TCAGAGTGGAG 800
 Db 147 SerGlnArgLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuL 166
 QY 801 ATGAAG 860
 Db 167 MetLysGlnPheLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 186
 QY 861 CGCAG 920

```

Db      187  LysGlnItyrSerLeuIysValThrIysLeuGlnHisAspAlaGlnAlaIysValGln 206
      1921  CTGAGTGAAGCCGAGCGGAGCGGAGCTCGAAGGCGCAAGAGCTCCAGGATCTGAC 980
      207  LeuIleGlnThrGlnIlysglnGlnGlnIleuGlnHisAspLeuSerIysAlaIa 226
      981  GAGCGATCTCGGCTCCAG---GAGTTCGCGAGAGGAGTCTGCGGCGCCAGACCAAGGTG 1037
      227  MetIysValIlysLeuGlnIlysglnPheArgIlySerIleAspAlaIalysLeuArgVal 246
      1038  CAGGTCTGGAAGGAAGAACCGGCTACCGAGCGGTGTGTCTGCTACGTCCGCGCCAGAGT 1097
      247  GlnValIleuGlnIlysglnGlnIAspSerIlysIleuAlaSerLeuSerIleGlnHis 266
      1098  GAGAAGCGACTGCGAGAGCTCGAAGCGGAGCTGAGCTCATGCGAGCGAGCAAGAGACAG 1157
      267  GlnIysArgIlaAsnGlnIleuGlnIleuGlnSerValAspHisMetIlyrGlnIlysgln 286
      1158  CTGCAAGAGCGGCTTCCGCGAGAGACCGAGCAAGAGCGGCTGAGAGCAAGAAATGAC 1217
      287  LeuGlnArgIlysLeuArgGlnGlnIAsnGlnIlysglnIleuAspAlaValIleIys 306
      1218  AAGCGGACGACCGCTCAAGAGAGCTGAGCTGAGCATGAGCAAGACAGAGAGAGATCTGT 1277
      307  ArgAspGlnGlnIlysglnIleuGlnIleuGlnIleuGlnIleuGlnIleuGlnIleu 325
      1278  AAGATTAAAGACGAGAGATGCGGCTTCCAGAGAGAGCGGCTGAGCGAGCAAGCGC 1337
      326  LysProIysAlaGlnIleuAspAlaCysAsnIleuIysArgIlysglnSerPheGly 345
      1338  TCTGTGGTCACTCGGACAGACAGCAAGATTAGAGAGAGAGAGAGAGAGAGAGAGAG 1397
      346  SerIle-----AspHisLeuGlnIlysglnIleuAspGlnIlysglnIleuAspGln 362
      1398  GAGATGGAAGAGTCTCTACAGCAGCGGCGGCGCTGAGAGAGCTGGGCGAGAGCTCCAC 1457
      363  GlnValGlnIlysglnIleuAsnGlnIlysglnIleuGlnIleuGlnIleuGlnIleu 382
      1458  AAGCGGAGCGCATCTCTGCGCAAGAGAGCGGCTGATGAGAGAGAGAGAGAGAGAGAG 1517
      383  LysArgGlnIleuAlaIleuValSerIlysglnIleuGlnIleuGlnIleuGlnIleu 402
      1518  AGCAAGCGCTGAGATCCAGCAGAGCGCTTCAACAGAGAGATCTGTGCTCCAGCGG 1577
      403  AsnIlysglnIleuArgSerSerGlnAlaIleuAsnThrAspSerIleuIlysglnSerThrArg 422
      1578  CTGAGAGCACTGAGAGAGAGAGCTGTCCGAGAGAGCGGCGAGCTGGCGAGAGCGGCG 1637
      423  LeuAsnLeuLeuGlnIleuGlnIleuSerIlysglnIleuAsnValGlnIleuGlnIleu 442
      1638  CAGAGCGACAGCAGATCCGCGGAGAGATCGACAGCGCTGCGCGAGAGAGAGAGAGAG 1697
      443  GlnGlnIlysglnIleuSerGlnIleuGlnIleuGlnIleuGlnIleuGlnIleu 462
      1698  CTCAGAGAGCGCTTGGAGATGAGCGGAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1757
      463  GlnIysArgIlysglnIleuAspValAspIlysglnIleuIysArgIlysglnIleuSerProGln 482
      1758  GAGAGCGGAGCGCTTCTCCAGTGGATGAGCGCATCGAGCGCTGAGAGAGAGAGAGAG 1817
      483  GlnGlnIleuIleuValLeuPheGlnIleuGlnIleuGlnIleuGlnIleuGlnIleu 502
      1818  TATTAAGATGAGCGCATCATGATCCGCGAGCGGCTGTCCGCGCTTCAAGCTGTGCTG 1877
      503  TyrArgAsnGlnSerIleGlnAsnArgIlysglnIleuAspIleuAspPheHisAsnLeu 522
      1878  TCCCATGCGAGATGAACCTCATGCGAGAGCTCATGCTCTCATCTCTCAAGAGACAG 1937
      523  SerArgGlnIleuAlaAsnValLeuGlnIlysglnIleuIleuIleuIleuIleuIleu 542
      1938  GCCCTCTCTGCAAGATTATTTGACAAAGGTGTGACGCTCCGAGAGAGAGAGAGAGAG 1997

```

```

Db      543  ThrIleuPheArgTyrPheAsnIlysglnValIleuAsnLeuArgIleuAlaGlnArgIysGln 562
      1998  CAGATTGCTTCTCGGAAGTGAAGATGCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 2057
      563  GlnIleuTyrAsnGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 582
      2058  CTGAGAGTGGCCCTGAGAGCGGAGCGCTCGAGATGAGACCGGACACTGAGACCTGAGAG 2117
      583  LeuGlnSerIleuAsnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 602
      2118  AAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2177
      603  LysGlnIleuIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 622
      2178  GGGTTACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2237
      623  GlyIleuMetIleuThrPheIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 642
      2238  GCGCGTTACATGTGATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2297
      643  TyrPheIlysglnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 660
      2298  GCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2357
      661  GlyGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 677
      2358  AAT-----GAGATGAGCTCCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2402
      678  AspIlyValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 697
      2403  CCCCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2447
      698  ArgProGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 717
      2448  TTGCTTCAGCGCTCGTTTACCC---TTGACCTGAGAA----- 2480
      718  LeuArgThrGlnProAsnProGlnIlysglnIleuIleuIleuIleuIleuIleuIleu 737
      2481  CGCTGAGAGCTGTGT-----GTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2516
      738  HisSerSerLeuAlaProProSerGlnIleuMetLeuGlnIleuGlnIleuIleuIleu 757
      2517  GAGGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2576
      758  AspAspAsnGlnPheThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 777
      2577  GAGGACAGCGCTGCGCTGGAACCTTGGGCTTGTCCAGAGAGAGAGAGAGAGAGAGAG 2636
      778  AsnValGlnIlysglnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 797
      2637  GCCAGCGCGGAGATGATGATGATCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2675
      798  IleSerIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 810

```

RESULT 3
 08BMB9 PRELIMINARY; PRT; 669 AA.
 AC 08BMB9;
 DT 01-MAR-2003 (TRENBLREL. 23, Created)
 DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Similar to hypothetical protein DKFZp44D0917.1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=C57BL/6J; TISSUE=Mesonephros;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RA "Analysis of the mouse transcriptome based on functional annotation of

Db	622	VallylsProVallylsLeuGlyCysArgGlyGluLeuArgGlnIleLeuSerIleLeuGluLeuSer	645
QY	2658	GNCCGGAAAACCCCTGTAGACCCCTCGGGCAGACCTCCCTTGAGAGGAACCTCGAG	2717
Db	646	LeuArgArgCysSerLeuGluAlaGlyGlyArgSerMetThrAlaAspSerLeuGluLysP	665
QY	2718	CCTGCTGAA	2726
Db	666	ProGluGlu	668
RESULT 4			
Q95UTC9		PRELIMINARY;	PRT; 676 AA.
ID	Q95UTC9		
CD	Q95UTC9		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
OS	Basic proline-rich protein.		
CS	Sus scrofa (Pig).		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.		
CC	NCBI_TaxID=9823;		
OX	[1]		
RN			
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Parotid gland;		
RA	Zhang Q., Szalay A.A., KyeYune-Nyombi E., Sands J.F., Oberg K.C.,		
RA	Tieche J.-M., Leonora J.,		
FT	"Cloning and expression of a novel proline-rich protein from porcine		
RT	parotid glands.";		
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.		
EMBL	AY035849; AA061383.1; -		
DR	Interpro: IPRO05404; KV3.3 channel.		
DR	Interpro: IPRO02965; P rich extensn.		
DR	PRINTS: PR01583; KV33CHANNEL		
DR	PRINTS: PR01217; PRICHEXTENSIN.		
Q	SEQUENCE 676 AA; 62297 MW; 3008BC41FDB1FC9 CRC64;		

Alignment Scores:			
Pred. No.:	1,9e-21	Length:	676
Score:	626.50	Matches:	258
Percent Similarity:	29.54%	Conservative:	12
Best Local Similarity:	28.23%	Mismatches:	331
Query Match:	10.96%	Indels:	313
DB:	6	Gaps:	47
US-10-032-596-6 (1-3121) x Q95JC9 (1-676)			
Cy	2706	CTCTCAAGGAGAGGTGTGCCCCGAGGCTTACAGAGGGGTTTTCGGACATCAATCATCC	2644
Db	43	ProProAlaGProProProProProGluInserInGlnly-----GluGlnHstGlnlyAlaG	60
Cy	2646	CCGAGCTGCTGTGTGCACTTCCGCGGGGCTTGAGCAAAAGGCCAAAGTTCCAGGGCA	2587
Db	61	ProArg-----ProProGlnly-----	65
Cy	2586	GGCTGCTCTACCCACAGAGAACACCCCGCCACACAGAGGCTCAGCGCTCCCGCTGCC	2527
Db	66	-----AspGlyProGluInGlnlyProAlaProProGlnlyAlaGluProProProGlnlyPro	83
Cy	2526	TCAGTTCTCTCGGGGAGACCCCTCTCTCTCAACACAGAGCTCGAGCGTTTCCAGGTCAAG	2466
Db	84	Pro--ProProGlnlyProProProPro-----GlyProAlaProProGlnlyAlaArg	99
Cy	2466	GTAACGAGAGGTTGAGACCAAGTCCCGCGCTCTCTCTCCGAGGAGCGGGGAGCCCTCAGTGA	2407
Db	100	-----ProProProGlnlyProProPro-----ProGlnlyProProProPro	112
Cy	2406	GGGGGAGACACCGACAGAAAGCTCTGGGTGGCCAGGTGAGACTCATCTTCACTTTCCAGAGGCT	2344
Db	113	GlyProAlaProProAlaArgProProProGlnly-----	124
Cy	2346	GTCGCGCTTCGAGACAGAGCTCTCTTCTCCCGACCCGCTGTGTGCGCTTACAGCGTTCA	2287

Db	125	-----ProProProProGlyProProProProGlyPro-----	133
Qy	2286	CACCGCGCAGCTCTGTTTTCAGTTCGCGGTTTTCACATGTAACGCGCAGTTCCTTCT	2257
Db	136	-----AlaProProGlyAlaAlaProProProGlyProProPro-----	146
Qy	2226	CCAGACCTTGAAATCCGGGCTCATATCGCTCTGCTGTCTACTAACCTTCACCGAGGT	2167
Db	149	-----ProGlyPro-----	151
Qy	2166	GGCTTCGACTGCTGCGACGAGCAGCTGATCTTTCCTCGTCTGCTCTGCTCAGAGG	2107
Db	152	-----ProProProGlyProAlaProProGlyAlaAlaProProProGlyProPro	168
Qy	2106	TCAGCTGGCGGTCATCTCCAGGCGCTCCGCTCCAGG-----	2068
Db	169	ProProGlyProProProProGlyProAlaProProGlyAlaAlaProProProGlyPro	188
Qy	2067	CCACCTCCGACCGCAGTAACCGACCTCTGCTGCTCCCTCCAGCGCATCTCCACTTCCAGA	2008
Db	189	ProProProGlyProProProProGlyProAlaProProGlyAlaAlaProProPro-----	207
Qy	2007	AGGCATCTGCTCTGTGTGCTGCTCTCTCGAGCGTCACCACTTGTCAAATACTTGC	1948
Db	208	-----GlyProProProProGlyProProProPro-----	217
Qy	1947	AGAGGAGGCGCTGCTCTGAGATGAGAGTAAGTAGAGCTTGCGCATGAGTTCATCT	1888
Db	218	-----GlyProAla	220
Qy	1887	CGCAGCTGGACAGCAACGAGGCTGAGGCCGGAAGACCCGCTGGCGGATGTGAGCTT	1828
Db	221	ProProGlyAla-----ArgProPro-----	227
Qy	1827	CATTCTTATACATGAGCAGCATCCAGGGGCTCGATGGCTCATCTCAATCTGAAACAGG	1768
Db	228	-----ProGlyProProProProGlyProProProProGlyProProProProGly	239
Qy	1767	TCCGCTCTCTCGGGGAGACAGAGACTCCCTCTGCTCAGCTTGGCCGTGATCTCAGCG	1708
Db	240	ProAlaProPro-----GlyAlaAlaProProProGly	250
Qy	1707	GCCTGTTAGCAGCAGATCTTCTCCAGC-----GCAGGCTGTGATCTCCGCGC	1657
Db	251	ProProProLeuGlyProProProProGlyProAlaProProGlyAlaAlaProProPro	270
Qy	1656	GGATCTGCTGCTGCTCGGGGCTGCGCTGCTCCGCGAGCT-----GCCGCTCTTCT	1606
Db	271	GlyProProProProGlyProProProProGlyProAlaProProGlyAlaAlaProPro	290
Qy	1605	CGCAGACGCTCTTCCGACGTGCTCCAGCGCGCTGAGCACTGCACAGATGCTCTGTGA	1546
Db	291	ProGlyProProProProGlyProProPro-----	300
Qy	1545	GGGCTGCTGGATCTCAGCGCGCTGCTCTTCACGCCCGCTCTTCTCTGCATCAGAGGCT	1486
Db	301	-----ProGlyProAlaProProGlyAlaAlaProProProGlyProProProProGlyPro	319
Qy	1485	-----CATTCTTGGCCAGATAGGCGCTCCGCTTGGAGACTCTCCGCCACACT	1438
Db	320	ProProProGlyProAlaProProGlyAlaAlaProProProGlyProProProProGly	339
Qy	1437	CTTCACGCGCCCGCC-----GCTGCTGTAACACTTTCATCTCTGCTCCAGCCACTTCT	1381
Db	340	ProProProProGlyProAlaProProGlyAlaAlaProProProGlyProProProPro	359
Qy	1380	TCTGCTCCCATCTTCTGCTGCTGCTTCCAGGCTGACACAGAGCCGTTGCTGCACATGC	1321
Db	360	GlyProProProProGlyProAlaProProGly-----	370
Qy	1320	GGCTCTTCTCTGGAAGCGCGCATCTTTCGCTTAACTTCAGAGATCTTCTGCTGT	1261
Db	371	-----AlaArgProProProGlyPro-----	377

```

QY 1260 GCTCATGCTTGAAGCTCCAGCTCTTGAAGCGGTGCTGCGCTTCTCATTTTCTGCTTCA 1201
Db 378 -----ProProPro-----GlyProProProProGlyProAlaProPro 390
QY 1200 GAGCGCGCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1141
Db 391 GlyAlaArgProProProGlyPro---ProProProGlyProProProProGlyProAla 409
QY 1140 GATAGAGCTGCACTG-----TCCGCTGAGACTCTCTGAGTCCGCTTCTCACTCTGG 1090
Db 410 ProProGlyAlaArgProProProProProProProProAlaArgProGlnGlnGly 429
QY 1089 CCGACAGTACACACCGCTCCGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1036
Db 430 Pro-----AlaProSerGlyAspGlyProGlyAspGlyProProProProAlaGlyPro 447
QY 1035 CCTGAGCTTGGGCGCGCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 976
Db 448 ProProProGlyProProSerProGlyProAlaProProGly-----AlaArgPro 464
QY 975 CAGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 916
Db 465 ProProProGlyProProProProProProProProProGlyProAlaProProGlyAla 484
QY 915 CCGGACCTGCTCTG-----CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 868
Db 485 ProProProGlyProProProProProProProProProGlyProAlaProProGlyAla 504
QY 867 GCTGGGCGTTGAGGCGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 808
Db 505 ArgProProProGlyProProProPro-----GlyPro 515
QY 807 CTTTCATGCGATGTTGATAGCAGCT---CCCGAGTCTCTCTCTCTCTCTCTCTCTCTCT 751
Db 516 ProProProGly-----ProAlaProProGlyAlaArgProProProGlyProPro 532
QY 750 GCGACTGTAGGCTGTGAGGAGGAGCTGCGGCGCTGAGCTGAGGCGCTGCTCTCTCT 691
Db 533 Pro-----ProGlyProProProProProProProProGlyProAlaPro--- 543
QY 690 CAAGTCTGTGAGCCCTGGAATGCTGTGATCAATCTCTCTCTCTCTCTCTCTCTCTCT 631
Db 544 -----ProGlyAlaArgProProProProProProProProProGlyPro 557
QY 630 TCTCTCTGTGAGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 571
Db 558 -----ProProProGlyProAlaPro----- 564
QY 570 TTCTGCGAGTGTAGATCCGCTGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 511
Db 565 -----ProGlyAlaArgProPro---ProGlyProProProProProProPro--- 579
QY 510 CAGCTGAAGAGCACTCTCCAGCTGTTCACTCACTCACTCACTCACTCACTCACTCTCC 451
Db 580 -----ProPro 581
QY 450 TGCATTGTGACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 391
Db 582 -----GlyProAlaArgProProGlyAlaArg----- 589
QY 390 CAGGCGGACCATGCCAGACATAGGCGGTGGGACCCCGAGGGGGGCTGTATAGAGTC 331
Db 590 -----ProProPro-----GlyProProPro----- 596
QY 330 GAGGCAAGAAGACCGGAGGAGGAGCATTCAGAGCGAGCGGCGGCGGCGGCTGCGC 271
Db 597 -----ProGlyProProPro---Pro 602
QY 270 GACACAGCTCTTAACCGGACCGCAGTTCACCATCTCTCTCTCTCTCTCTCTCTCTCTCT 211
Db 603 GlyProAlaArgProGlyAlaArgProProProProPro-----ProProGly 619

```

```

QY 210 CGCTCTGAGTTGTACTGCTCA-----TGAGCTCTCTCCAGCGCAGCAAGATCTC 157
Db 620 ProProProProProGlyProAlaProProGlyAlaArgProProProGlyProProProPro 639
QY 156 GGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 97
Db 640 GlyProProProProProGlyProAlaProProGlyAlaArgProProProGlyProProPro 659
QY 96 CAGCTGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 55
Db 660 ProProProGlyProSerProProProProProProProProProProProProPro 673

RESULT 5
Q9BIT7 PRELIMINARY; PRT; 1953 AA.
ID Q9BIT7
AC Q9BIT7;
DT 01-JUN-2001 (TRMBLrel. 17, Created)
DT 01-JUN-2001 (TRMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TRMBLrel. 22, Last annotation update)
DE Major ampullate spideroin 2-like protein (fragment).
OS Nephila madagascariensis.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OC NCBI_TaxID=115969;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gately J., Hayashi C., Mottick D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
RT Fibroin Sequences."
RL Science 291:2603-2605 (2001).
DR EMBL: AF350276; AKX0605.1; -.
FT NON_TER 1
FT NON_TER 1953
SQ SEQUENCE 1953 AA; 159383 MM; C82B4DACD043C3BD CRC64;

Alignment Scores:
Pred. No.: 4,096-20 Length: 1953
Score: 598.00 Matches: 309
Percent Similarity: 32.35% Conservative: 41
Best Local Similarity: 28.56% Mismatches: 409
Query Match: 10.42% Indels: 323
DB: Gaps: 51

US-10-032-996-6 (1-3121) x Q9BIT7 (1-1953)
QY 53 CGAGGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 109
Db 120 ArgGlyProGlyGlyTyrgLProGlyGlnGlnGlyProGlyGlyProGlyAlaAla 139
QY 110 GCTGCT-----GACCTGACAGAACGAGTGGC 136
Db 140 AlaAlaAlaGlySerGlnGlyTyrgLProGlyGlnGlnGlyProArgGlyProGlyAla 159
QY 137 GCGGCTGAGGA-----GGAAGACCGAGACTTCTGCG 169
Db 160 AlaAlaAlaGlyProGlyGlyTyrgLProGlyGlnGlnGlyAlaSerAlaAlaAlaSer 179
QY 170 TGCCTGAGGAGCGC-----CATGAGCAGTACAACTCAGAGCGACCGCGCT 217
Db 180 AlaAlaAlaGlyArgGlyProGlyGlyTyrgLProGlyGlnGlnGlyProGlyGlyPro 199
QY 218 GCGTACGACGAGGAGGAGATGTGCA-----ACTGGG 250
Db 200 SerAlaAlaAlaAlaGlyProGlyGlyTyrgLProGlyGlnGlnGlyProSerAlaAla 219
QY 251 GCTGCGGTTAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 292
Db 220 AlaAlaAlaAlaAlaGlySerGlyProGlyGlyTyrgLProGlyGlnGlnGlyProGly 239
QY 293 GCTCTGATAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 352
Db 240 GlyProGly-----AlaAlaAlaAlaAlaAlaGlyProGly 251

```

QY	353	GGGTGCCACGCCCATGTGCTGGGCGATGTCGCCGCTGCTGCTCCTCGAGATGAAGT	412
Db	252	GLYTYRGLYProGlyGlnGlnGlyProGlyYAlaAlaAlaAlaA----	265
QY	413	TGAGCTCTGACAGAGGGAGAGCGAGCTGACAAATGCGAGGAGGCTGAGCTGATGCT	472
Db	266	-----AlaAlaGlyAArgGly-----ProGlyGlyTYrGly-----	275
QY	473	GACTAGCTGAACAG-----CTGGGAAGTGGCTCTTACGCTGCTTCAAGAGAGMAA	526
Db	276	---ProGlyGlnGlnGlnGlyProGlyGlyProGlyYAlaAlaAlaAlaAlaAlaGlyAArg	294
QY	527	GGAGAGAGAGAGCGCGCCAGCGAGCCTTAACCTGCGCAGAAATGAGATCAGCACTG	586
Db	295	GlyProGlyGlyTYrGly-----ProGlyGln-----	303
QY	587	CAGTCAGAGGGGGGGGCGACCGCCAGAGGAGCTCCAGAGAGAGMAAGGCCAGACTTG	646
Db	304	-----GlnGlyProGlyGlyProGlyYAlaAlaAlaAlaAlaAlaGlyProGlyTYr	321
QY	647	CCTTAGGAGATTGATGACAGCATTTCCAGGGTCCAGAGCAGTTGATGGAGCAGAGCCCG	706
Db	322	-----GlyProGly-----GlyTYrGlyPro	328
QY	707	AGTTACGGCGCCGCGAGCTCCCTCTGCGACAGCTCAGAGTGGCGGCTGGCCAGGCCCA	766
Db	329	GlyGlnGlnGlyProGlyGlyProGlyYAlaAlaAlaAlaAlaAlaGlyAArgGlyPro	348
QY	767	GCA-----GAAGATCCGGAGGCTGGCTATCAATCCGATCAGMAAGAGAGCT	814
Db	349	GLYLYTYRGLYProGlyGlnGlnGlyProGlyGlnGlnGlyProGlyGlyA	368
QY	815	TATGGCGAGCT-----GCTCCGACAGMAAGGCGACTCA	850
Db	369	AlaAlaAlaAlaAlaGlyAArgGlyProGlyGlyTYrGlyProGlyGlnGlnGlyProGly	388
QY	851	GGCCTCGAAGCCCGACAGCAGCGCTATCCGGAGCTGAGACAGAGCAGACAGT	910
Db	389	GlyProGlyYAlaAlaAlaAlaAlaAlaGlyProGlyGlyTYrGlyProGlyGlnGlnGly	408
QY	911	GCGGCGCAGCTGAGTGAAGGCGAGAGCGAGCTGCG-----	946
Db	409	ProGlyYAlaAla-----AlaAlaAlaAlaAlaAlaGlyAArgGlyProGlyGlyTYrGly	426
QY	947	---GAGACTCGAGGCGAAGAGCTTCAGAGATGCTGAGCGAGCTTCGAGCT-----	994
Db	427	ProGlyGlnGlnGlnGlyProGlyGlyProGlyYAlaAlaAlaAlaAlaAlaAlaGlyAArgGly	446
QY	995	-----CCAGGATTCGAGAGAGGCTGCTGCGCGCCAGAGCGAGCTGCA	1039
Db	447	ProGlyGlyTYrGlyProGlyGlnGlnGlyProGlyGlnGlnGlyPro--GlyGlyA	465
QY	1040	GGTGCTGAAGAGAAAGACAGCTTACGAGCGGCTGATCTCACTGCGCCACAGTGA	1099
Db	466	GlyYAlaAlaAlaAlaAlaAlaAlaGlyAArgGlyProGlyGlyTYrGlyProGlyGlnGlnGly	485
QY	1100	GAGCGACTGCGAGAGCTCGAGCGGAGAGCTGCACCT-----	1135
Db	486	ProGlyGlyProGlyYAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlyTYrGlyProGlyGln	505
QY	1136	CATGCGGACAGCAGCAGGACAGCTTCAGAGCGGCTTCGCGAGAGACGGA-----	1186
Db	506	GlnGlyProGlyYAlaAlaAlaAlaAlaAlaAlaAlaGlyAArgGlyProGlyGlyTYrGly	525
QY	1187	---GCAAGAGCGCGCTGAGAGCGAAGATAGCAAGCGGACACCGGCTCAAGAGAGT	1243
Db	526	ProGlyGlnGlnGlyProGlyGlyAArgGlyProGlyYAlaAlaAlaAlaAlaAlaGlyAArgGlyPro	545
QY	1244	GAGAGCTGAAGATGAGCAACAGCAGAGATCTCTGAAGATTAAAGCGAAAGAGTGC	1303
Db	546	Gly-----GlyTYrGlyProGly	551

QY	1304	CTTCCAGAGAAAGAGCGCAGTGGCAGCAACGGCTCTGTGCTCAAGCTCTGGAAACGACGA	1363
Db	552	-----GlnGlnGlyProGlyGlyProGlyAlaAlaAla	562
QY	1364	GAAGATTGAGAGCAGAAAGATGGCTGACCGACGA	1405
Db	563	-----AlaAlaAlaGlyProGlyGlyTyTyrGlyProGlyGlnGlnGly	576
QY	1406	GAAGGTGCTACAGCAGCGCGCGCGCTGGA	1447
Db	577	ThGlyAlaAlaAlaAlaAlaAlaAlaAlaGlySerGlyAlaGlyGlyTyTyrGlyProGlyGln	596
QY	1448	-----GGAGCTCCACAGCG	1474
Db	597	GlnGlyProGlyGlyProGlyAlaAlaAlaAlaAlaAlaGlyProGlyGlyTyTyrGlyPro	616
QY	1475	GGCGAAGAGAGAGCGCGCTGATGCGAGGAGAGCGGGGCTGGAGAGCAGACCGCTGAGATC	1534
Db	617	GlyGlnGlnGlnGlyProGlyAlaAlaAlaAlaAlaAlaGlySerGlyProGlyGlyTyTyr	636
QY	1535	CAGCCAGCGCCCTCAA	1582
Db	637	GlyProGlyGlnGlnGlnGlyProGlyGlySerSerAlaAlaAlaAlaAlaAlaGlyProGly	656
QY	1583	-----GCACCTGGAGAAAGAGCGTGTCCGAGAAAGCGGAGCGAGCTGGCGGAGGAGCGC	1636
Db	657	ArgTyrGlyProGlyGlnGlnGlnGlyProGlyAlaAlaAlaAlaAlaAlaSerAlaGlyArgGly	676
QY	1637	CCA	1648
Db	677	ProGlyGlyTyTyrGlyProGlyGlnGlnGlnGlyProGlyGlyProGlyAlaAlaAlaAlaAla	696
QY	1649	GCAGATCGCGGGGAGATCGACAGCTCGCGCCGACGAGAGAGAACTCGCTCTCAACGACAGC	1708
Db	697	AlaGlyProGlyGlyTyTyrGlyProGlyGlnGlnGlnGlyProGlyAlaAlaAlaAlaAlaAla	716
QY	1709	-----CCTGGAGATCGACGCGCAAGCTGGAGCAGGGGAGTCTGCTGCCCGCA	1756
Db	717	AlaGlySerGlyProGlyGlyTyTyrGlyProGlyGlnGlnGlnGlyPro	731
QY	1757	GGAGAGCGGAGAGCTGTTCCAGTTGGATGAGAGGCGATCGAGCGCTCGATGCTGCATTGA	1816
Db	732	GlyGlyProGlyAlaAlaAlaAlaAlaAlaAlaAlaGlyArgGlyProGly	747
QY	1817	GTATTAAGATGAGGCATCACATGCCGCCAGCGGGTGTCTGGGCTCAGCCTCGTTGCT	1876
Db	748	-----GlyTyTyrGlnGlnGlyGlnGlnGlnGlyProGlyGlyProGlyAlaAlaAlaAla	763
QY	1877	GTCCAGTGCAGATGAACTCAACGAGCGCAAGCTCAGTCACTCATCTCCAGAACAG	1936
Db	764	AlaAlaAlaGlyProGlyGlyTyTyrGlyProGlyGln	776
QY	1937	AGCCCTCTCTCCAAATATTTTGACAAAGGTGTGACGCTCCGAGAGAGACACACACGA	1996
Db	777	GlyProGlyAlaAlaAlaAlaAlaAlaAlaAlaGlySerGlyProGlyGlyTyTyrGlyProGly	796
QY	1997	GCAGATTGCTTCTCGAACTGAGAGATGACAGCTGGAGAGACGACAGAGCTGTACTATG	2056
Db	797	GlnGln	811
QY	2057	GCTGGAGGTGGCCCTTGGAGCGCAGCCGCTTGGAGATGAGACCGCTGACCTGACCTGACCA	2116
Db	812	AlaGlyArgGlyProGlyGlyTyTyrGlyProGlyGlnGlnGlnGlyProGlyGlyProGlyAla	831
QY	2117	GAAGAGCAGACGACA	2143
Db	832	AlaAlaAlaAlaAlaGlyProGlyGlyTyTyrGlyProGlyGlnGlnGlnGlyProGlyAlaAla	851
QY	2144	GCTCTCGACACAGACTCGAGACCACTCGGTGAAGGTTTACAGACAGACGAGAGGACGACTA	2203
Db	852	AlaAlaAlaAlaGlyArgGlyProGlyGlyTyTyrGlyProGlyGlnGln	868
QY	2204	TGAGGCCCGGATTTCAGC	2230

```

Db      869  ---GlyProGlyGlySerGlyAlaAlaAlaAlaAlaAlaGlyArgGlyProGlyGlyTyr 887
QY      2231  GGAAGTGGGCGGTTACATGTGTGAATAACAGAACTAAACAGAACTCGCGGTGTGA 2250
Db      888  GlyProGlyGlyGlyGlyProGlyGlyProGlyAla---AlaAlaAlaAlaAlaGly 906
QY      2291  CGCTGTAGCCACAGACAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2350
Db      907  ArgGlyProGlyGlyTyrGlyProGlyGlyGlyGlyProGlyGlyGlyGlyGly 923
QY      2351  TCCTGGAAATGATGATGAGTCTCCACCTGGACCCGAGACTTCTGTGCTGCTCC 2404
Db      924  -----ProGlyGlySerGlyAlaAlaAlaAlaAlaAlaGlyArg 936
QY      2405  -----CTCACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2434
Db      937  GlyProGlyGlyTyrGlyProGlyGlyGlyGlyGlyGlyProGlyGlyAlaAlaAla 956
QY      2435  -----GGAAGAGCGGAGCTGTGCTCCAGCTCCCTTACCTGACTGAAAGCTCGAG 2488
Db      957  AlaAlaGlyProGlyGlyTyrGlyPro----- 965
QY      2489  CCTGTGTGTGAGAGAGAGGAGTCCCGGAGAACTAGAGAGAGAGAGAGAGAGAGAG 2548
Db      966  -----GlyGlyGlyGlyProGlyAlaAlaAlaAlaAlaAlaGlySerGly 981
QY      2549  CCTGTGTGTGAGAGAGGAGTCCCGGAGAACTAGAGAGAGAGAGAGAGAGAGAGAG 2605
Db      982  ProGlyGlyTyrGlyProGlyGlyGlyGlyProGlyGlyProGlyAlaAlaAlaAla 1001
QY      2606  TTTGTCCAAAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2665
Db      1002  AlaGlyArgGlyProGlyGlyTyrGlyProGlyGlyGlyGly----- 1015
QY      2666  AAACCCCTGTAAAGCCTCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2725
Db      1016  -----ProGlyGlyProGlyAlaAlaAlaAlaAlaAlaGlyArgGlyProGly 1033
QY      2726  AAGGGGAGAGCTGCTTTGCTTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2785
Db      1033  YTYGly-----ProGlyGlyGlyGlyProGlyGlyGlyGly 1046
QY      2786  AGGC 2789
Db      1046  oGly 1047

RESULT 6
QY1B4      PRELIMINARY;      PRT: 1450 AA.
ID  QY1B4
AC  QY1B4;
DT  01-MAY-1999 (TREMBlrel. 10, Created)
DT  01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT  01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE  Alpha 1 type I collagen.
OS  Cynops pyrrhogaster (Japanese common newt).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Amphibia; Batrachia; Caudata; Salamandridae; Cynops.
NC  NCBI_TaxID=8330;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Regenerate forelimbs;
RA  MEDLINE=99407244; PubMed=10474166;
RA  Asahina K., Obara M., Yoshizato K.;
RT  "Expression of genes of type I and type II collagen in the formation
RT  and development of the blastema of regenerating newt limb.";
RL  Dev. Dyn. 216:59-71(1999).
DR  EMBL: AB015438; BAA36973.1;
DR  InterPro: IPR000087; Collagen.
DR  InterPro: IPR000885; FIB_Collagen_C.
DR  InterPro: IPR001007; VWF_C.
DR  Pfam: PF01410; COLFI; 1.
DR  Pfam: PF01391; Collagen; 18.

```

```

DR  Prodcm; PD000007; Collagen; 1.
DR  Prodcm; PD002078; FIB_collagen_C; 1.
DR  SMART; SM00038; COLFI; 1.
DR  SMART; SM00214; VWC; 1.
DR  PROSITE; PS01206; VWC; 1.
KM  Collagen.
SQ  SEQUENCE 1450 AA; 137564 MW; AF8A74841B87B7C CRC64;

Alignment Scores:
Pred. No.:      2,19e-18      Length:      1450
Score:          560.50      Matches:      291
Percent Similarity: 31.58%      Conservative: 63
Best Local Similarity: 25.96%      Mismatches: 416
Query Match:    9.81%      Indels:      351
DB:              13      Gaps:      56

US-10-032-996-6 (1-3121) x QY1B4 (1-1450)
QY      2770  GGTAGAGACTGCCCCCTTACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2711
Db      216  GlyAspAspGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 226
QY      2710  TCTCCCTCCAGAGAGAGG--TCTGCCCGAGAGGCTTACAGAGGCTTTTCCGACATCA 2654
Db      227  ---ProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 245
QY      2653  ATCATCCCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2594
Db      246  GlyLeuProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 264
QY      2593  CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2534
Db      265  AspAsnGlyProAlaGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 284
QY      2533  CGCTGCTCAATTCCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2495
Db      284  OGlyGlyAlaGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 304
QY      2494  CACAGGCTCGAGAGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2435
Db      304  AlaGly-----AlaArgGlyAsnAspGlySerProGlyAlaAlaGlyPr 319
QY      2434  TCCCGGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2416
Db      319  OProGlyProThrGlyProThrGlyProProGlyPheProGlyAlaValGlyAlaGly 339
QY      2415  -----CCTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2366
Db      339  YAspAlaGlyProIn-----GlySerArgGlySerGlyGlyProGlyGlyAlaArg 356
QY      2365  TCTTCATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2312
Db      356  G-----GlyGlyProGlyAlaProGlyProAlaGlyAlaAlaGlyProSerGlyAs 373
QY      2311  CCCCTGCTGTGGCTTACAGAGCTTCAACCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2266
Db      373  nProGlyThrAspGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 393
QY      2265  -----GTTCCTGTTATTCACATGTAACGGCCCA-----GTTC 2231
Db      393  YAlaProGlyPheProGlyAlaArgGlyAlaProGlyProGlyGlyProAlaGlyAla 413
QY      2230  TTCTCCAGAGCTTGAATCCGGGCTCACTACTGCTCTGCTGTCTGTAACCTTCACCG 2171
Db      413  oGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 425
QY      2170  AGGTGTCTGACTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2111
Db      426  -----GlyAsnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 440
QY      2110  AGGTCAAGCTGAG-----GGTCAATCT----- 2089
Db      440  aGlyValGlyGlyProGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 460

```

QY 2088 -CCAGGCGCTGCCCTCCA-----GGGCGCACTCCAGCCAG 2054
 Db 460 uProGlyProAlaGlyProProGlyProAlaGlyGluArgGlyGlyProGlySerArgGly 480
 QY 2053 TACACGAGCCTCT-----GCTGCTCTCCAGCTGCATCTCCAGTTCCGAGAAGCCATC 2000
 Db 480 yPheProGlySerAspGlyAlaSerGlyProGlyGlyAlaPro---GlyGluArgGlySer 499
 QY 1999 TGTCTGTGCTGCTGCTCTCTCCAGAGCTCACCACCTTCTCAAAATCTTCGAGAGAG 1940
 Db 499 fValGlyProAlaGlyProGlySerThrGlyGluSerGlyArgProGlyGluProGly 519
 QY 1939 GCTGTGCTCTGAGAG---ATGAGAGTACCTGAGCTTGGCCATGAGATTCACTTCAGCAC 1883
 Db 519 yLeuProGlyAlaGlySerGlyLeuThrGlySerProGlySerProGlyProAlaGlySerTh 539
 QY 1882 TGGACACGACACGAGGCTGAGAGCCCGAGACCCGCTGGCGGATGTGATGCTCATTC 1823
 Db 539 rGly-----ProAlaGlyAlaAlaGly----- 546
 QY 1822 TTATACCTCATGACGATCCAGAGGCTCGATGCTCATCCAACTGAAACAGCCGCCG 1763
 Db 547 -----GlnAspGlyHisProGlyPro-----ProGlyProSerGlyAlaArgGlyGly 562
 QY 1762 TCCT-----CCTCGGGGAGACGACGACTCCCTGCTGCTCAGCTTG 1724
 Db 562 nSerGlyAlaMetGlyPheProGlyProGlySerGlyAlaAlaGlyGluProGlyLysSerGly 582
 QY 1723 CCGTGCAGTCTCCAGCGCTGCTTAGACGAGAGTCTCTCCGCGGAGAGCTGTGATC 1664
 Db 582 yGluArg-----GlyAlaAlaGly----- 588
 QY 1663 TCCCGCGGATCTGCTGCTGCTGCTGCGCGCTGCCCTCCGCGAGCTGCCCTTCTCG 1604
 Db 588 yProProGlyAlaThrGlyAlaProGlyLysAspGlyGlyAlaGlyAlaGlnGly----- 606
 QY 1603 GACAGCTCTCTCCAGAGCTCCAGCGCGCTGACACCTGCAAGATGCTCTGTTGAGG 1544
 Db 607 ---ProProGlyProSerGlyProSerGlyGluArgGlyGluGlnGlyProAlaGlySer 625
 QY 1543 GCTTGCTGATCTCAGGCGCTTGTCTCCAGCCCGCTCTCTTCCTGCATCAGAGGCTCC 1484
 Db 625 rProGlyPheGlnGlyLeuProGlySerProGlyProAlaGlyGluAla---GlyLysPr 644
 QY 1483 TTCTTGCCGAGAGTGGCTCCCGCTGTGAGAGCTCTCCCGAGCTCTCTCA----- 1432
 Db 644 oGlyGluGlnGlyAlaProGlyAspAlaGlyGlyProGlyProSerGlyProArgGlyGly 664
 QY 1431 -----GCGCGCGCGCTGTGTACACCTTCTCC 1403
 Db 664 uArgGlyPheProGlyGlyAlaArgGlyGlyGlnGlyProAlaGlyAlaGlnGlyProArgGly 684
 QY 1402 ATCTTCGCTCAGGCACTTCTTCTGCTCTCAATCTTCTGCTGCTTCTCCAGGCTGACC 1343
 Db 684 ySerProGlySerProGlyAsnAspGlyAlaGlyGlyGlyAlaGlyAlaAlaGlyAlaPr 704
 QY 1342 ACAGAGCGGTTGCTGCCACTGCGGCTTCTCTCGAAGGCG-----GATCTCTTCC 1289
 Db 704 oGlyGlyArg-----GlyProProGlyLeuGlnGlyMetPr 716
 QY 1288 GTCTTAATCTTCAGAGTCTTGTGCTGTGCTCATGCTTCAGCTCCAGCTCTTGAGCGCG 1229
 Db 716 oGly---GluArgGlySerAlaGlyMetProGlyAlaGlySerGlyAsp-----ArgGly 732
 QY 1228 TGTGCGCGCTGCTCA----- 1213
 Db 732 yAspAlaGlyThrGlyGlyAlaAspGlyAlaProGlyLysAspGlyAlaArgGlyLeuTh 752
 QY 1212 -----TTCTGCTCCAGAGCGCGCTTGTGCTCTCTCTCCGAGAGCGCGCTCTGC 1160
 Db 752 rGlyProThrGlyProProGlyProSerGlyAlaProGlyAspLysGlyGlnGlyPr 772

QY 1159 AGCTGTCCCTGCTGCTGCCGATGAGCTGCAGCTTCGCTGAGCTCCTGATGCTTC 1100
 Db 772 oSerGlyProAla----- 776
 QY 1099 TCACTCTGGGCGGACGATGACACACAGCGGCTCCCTAGCTGCTTCTCTTCCAGCAC 1040
 Db 777 -----GlyProThrGlyAlaArgGlySerProGlyGluArgGlyGluProGlyAlaPr 794
 QY 1039 TGCAGCTGCTGAGCGCCGACGACCCCTCTCGCGAACTCTCGAGGCC----- 991
 Db 794 oGlyProAlaGly-----IleCysGlyProProGlyAlaAspGlyGluPr 809
 QY 990 -----GAAACGCTCGCCAGCATCT-----GAGCTCTCTGGCC 956
 Db 809 oGlyAlaGlyGlyGluSerGlyAspAlaGlyProGlySerGlyAspAlaGlyAlaProGlyPr 829
 QY 955 TCGAGCTCCGCGAGCTGCTTGGCCTT-----CACTCAGCTCGGCGCGACCT----- 907
 Db 829 oAlaGlyProThrGlyAlaProGlyProAlaGlyAlaThrGlyPheProGlyAlaAlaGlyArgGly 849
 QY 906 -----GCTCTGCTCTGCTCCAGCTCCCGATACGCTGCTGCTGCGGCTTC 857
 Db 849 yThrArgGlyAlaAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArgGly 869
 QY 856 AGGCGCTGAGCTGCTTCTCTGCGGACGAGCTGCGCAATAGCTCTCTTCACTGCGG 797
 Db 869 uGlyProProGlyPro-----SerGlySerAlaGlyProProGlyProProGlyProGly 887
 QY 796 ATGTGATGACAGCTCCCGATCTTCTGCTGCGCTGCGCGAGCGGCACTCTAGAGCT 737
 Db 887 yGlyGlyGluGlyAlaGlySerGlySerArgGlyGluThrGlyProAlaGly-----ArgSe 905
 QY 736 GTGACAGGAGGACCTGCGCGCTGGAATCTCGGCGCTTCTCCACCACTGCTCGGAC 677
 Db 905 rGlyGluProGlyProAlaGlyPro----- 913
 QY 676 CTTGGAATGCTGATCATCACTCTCCAGGCAAGCTTGGGCGCTTCTCT-----CT 623
 Db 914 -----ProGlyProSerGlyGluGlySerProGlySerAspGlyPr 928
 QY 622 GCGAGACTCCCTGGGCGCTGCGCGCTCTGACTGCAATGCTATCTTATCTTCTGCGC 563
 Db 928 oAlaGlyAlaProGlyLeuProGlyProGlnGly-----IleAl 941
 QY 562 AGGTGATAGGCTCGGCGCTGCTCT----- 535
 Db 941 aGlyGluArgGlyAlaValGlyLeuProGlyGluArgGlyGluArgGlyPheSerGlyLe 961
 QY 534 -CTCTCTCTCTT-----CTCTCTGAAGCTGAAAGAC-----CACTT 494
 Db 961 uProGlyProAlaGlyGluProGlyLysGlnGlyProSerGlyProAsnGlyGluArgGly 981
 QY 493 CCGAGCTGTTACCTCAGTACGACCACTGAGCTCCAGCTCCCTGCGCATTTCTCACTGC 434
 Db 981 yProPro---GlyProSerGlyProProGlyLeuGlyGlyPro-----ProGly 996
 QY 433 TCTCCCTCTGCTCAGAGCACTCATCTTCCAGGGA-----GCGAGGACGCGGACCC 380
 Db 996 yGluProGlyArgGluGlySerProGlySerGlnGlyAlaProGlyArgAspGlySerPr 1016
 QY 379 ATGCCAGACATGCGGCTGGGAC-----CCCGCGAGG----- 346
 Db 1016 oGlyProLysGlyAspArgGlyGluGlnGlyProSerGlyProProGlyAlaProGlyAla 1036
 QY 345 -----GGCTGTATAGGTGAGGACCAAGGACCCGAGAGC 308
 Db 1036 aProGlyAlaProGlyProValGly-----ProAlaGlyLysAsnGlyAspAr 1052
 QY 307 AGGCAATTCAGAGAGCCGACGCGCGCGCTGCGGACGACGACCTTAACGCGAGCGCG 248
 Db 1052 gGlyGluThrGlyProAlaGlyProAlaGlyProAlaGlyProSerGlyValaArgGlyAla 1072
 QY 247 AGTT----- 244

Db 533 Pro-----GlySer--- 535
 QY 1908 GCTTGGCCATGAGTTTCATCTCCGACTGTGGACAGCAGAGCTAGGCCCGGAAAGACACC 1849
 Db 536 -----ProGlyProAspGlyLysThrGly-----ProProGlyPro 547
 QY 1848 GCTGGCGGACATGATGAGCCTCATCTTATACATATGACAGATCCAGGCGCTGATG 1789
 Db 548 AlaGly-----GlnAspGlyAspProGlyProAlaGly 558
 QY 1788 CTT-----CATCACTGGAACAGCGCTCCGCTCT---CTTCGGGGAGACAGA 1744
 Db 559 ProProGlyAlaAspGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyThrAla 578
 QY 1743 GACTCCCGCTGCTGAGCTTGGCCGCTGATCTCCAGGCGCTGCTGAGCAGAGCTCTCT 1684
 Db 579 GlyGluProGlyLysAlaGlyGluArg----- 587
 QY 1683 CTGGGCGACAGCTGTGATCTCCCGGAGATCTGCTGCTGCTGGCGCTGCCCTGCC 1624
 Db 588 -----GlyValProGlyProProGlyAlaValAlaGlyProAlaGlyLysAspGlyGlu 604
 QY 1623 GCACTGCGCGCTCTTCTGCGACAGCTCCCTTCGACGAGTGCAGCGCGCTGACACTGC 1564
 Db 605 AlaGlyAlaGlnGly-----AlaProGlyProAlaGlyProAlaGlyAlaArgGly 621
 QY 1563 GCAAGATGCTCTGCTGAGGAGCTGTGCTGATCTCAAGGCGCTTCTCTCAAGCCCGTCT 1504
 Db 622 GlnGlnGlyProAlaGlySerProGlyPhe---GlnGlyLeuProGlyProAlaGlyPro 640
 QY 1503 TCTCTCATCAGAGGCTCTCTTCTGCGCAGATGAGCTCCGCTGTGTGAGCTCTCTCC 1444
 Db 641 ProGlyGlnAlaGlyLysProGlyGlnGlnGlyValProGlyAspLeuGlyAlaProGly 660
 QY 1443 CCAAGT-----CTTCCA 1432
 Db 661 ProSerGlyAlaArgGlyGluArgGlyPheProGlyGluArgGlyValGlnGlyProPro 680
 QY 1431 GCGCGCGCGCTGTGAGCCTCTTCTGCGCAGATCTCTGCTGCTGCTGCTGCTCTCT 1372
 Db 681 GlyProAlaGlyProArgGlyAspAsnGlyAlaProGly-----AsnAspGlyAla 697
 QY 1371 CAATCTCTGCTCTGCTTCTCAAGGCTGACACAGAGCGCTGCTGCTGCTGCTGCTCTCC 1312
 Db 698 LysGlyAspThrGlyAlaProGlyAlaPro-----GlySer 709
 QY 1311 TCTGAAAGCGG-----CGATCTCTTCTGCTTAACTTTCAGGATCTCTGCTGTGCT 1258
 Db 710 GlnGlyAlaProGlyLeuGlnGlyMetPro-----GlyGluArgGlyAlaAla 725
 QY 1257 CATGCTTCAAGCTCCAGCTCTCTGACGCGGTGCTGCGCTGCTCA----- 1213
 Db 726 GlyLeuProGlyProLysGlyAspArgGlyAspAlaGlyProLysGlyAlaAspGlySer 745
 QY 1212 -----TTTCTGCTTCCAGCGCGCTTCT 1189
 Db 746 ProGlyLysAspGlyValArgGlyLeuThrGlyProIleGlyProProGlyProAlaGly 765
 QY 1188 GCTCCGCTCTCTCCGGAAGCGCTCTGACGCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1129
 Db 766 AlaProGlyAspLysGlyGluAlaGlyProSerGlyProAlaGlyProThr---GlyAla 784
 QY 1128 GCTTCCGCTCGAGCTCTGACGCTTCTCACTCTGAGCGCGACAGTGAACCAAGCGCT 1069
 Db 785 ArgGly-----AlaProGlyAspArgGlyGluAlaGlyPro----- 796
 QY 1068 CCGTAGCTCTCTCTCTCTTCCAGCAGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCT 1009
 Db 797 -----ProGlyProAlaGlyPheAlaGlyProProGlyAlaAspGlyGlnProGlyAla 814
 QY 1008 TGCGGAATCTCTGAGCGGACAGCGCTGACAGAGCTCTGAGCTCTGCTGCTGCTGCTG 949
 Db 815 LysGlyGluProGlyAspThrGlyValLysGlyAspAlaGlyProProGlyProAlaGly 834

QY 948 CCGGAGCTGCTGAGCTT-----CACTCAGCTGCGCGGACCT----- 907
 Db 835 ProAlaGlyProProGlyProIleGlyAsnValAlaGlyAlaProGlyProLysGlySerArg 854
 QY 906 ---GCTTGCCTCTCTCTCTCCAGCTCCCGATACCTGCTGTGCTGCTGCTGCTGCT 850
 Db 855 GlyAlaAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyPro 874
 QY 849 GAGCTGCTCTTCTGCGGACAGCTGCGCAATAGCTCTCTTCAATGCGAGATGTGA 790
 Db 875 Pro-----GlyProSerGlyAsnAlaGlyProProGlyPro----- 886
 QY 789 TAGCCAGCTCC---GATCTTCTGCTGAGCTGCGCTGCGGCGGCGCACTGAGGCTGTG 733
 Db 887 ---ProGlyProValGlyLysGlnGlyLysGlyProArgGly----- 900
 QY 732 CAGGGGAGACCTGGGCGG-----CTGAACCTGGGCGCTGCTCCACCACTGCTG 679
 Db 901 GlnThrGlyProAlaGlyArgProGlyGluValGlyPro----- 913
 QY 678 ACCCTGAATGCTGATCCACTCTCTCAAGCAAGCT-----CTGGGC 634
 Db 914 -----ProGlyProProGlyProAlaGlyGlnGlyLysGlySerProGly 927
 QY 633 CTTTCTCTTCTGACAGACTCTCTGCGCGCTGCGCGCGCTCTGACTGCACTGTGATCC 574
 Db 928 AlaAspGlyProAlaGlySerProGlyLysProGlyProGlnGly----- 942
 QY 573 TATTTCTGCGAGGTGTAGTCCGCTGCGCGGCTCTCTCT----- 532
 Db 943 -----IleAlaGlyGlnArgGlyValAlaGlyLeuProGlyGlnArgGlyLysArgGly 960
 QY 531 ---CTCTCTCTTCT-----CTTGAAGCAGCTGAAGCACTCTCCA 490
 Db 961 PheProGlyLeuProGlyProSerGlyGluProGlyLysGln----- 974
 QY 489 GCGTGTTCAGCTCAGTCAAGCACT---CAGCTCCAGCTCTGCTGCTGCTGCTGCT--- 436
 Db 975 ---GlyProSerGlyAlaSerGlyGlnArgGlyProProGlyProMetGlyProPro 992
 QY 435 -----GCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 412
 Db 993 GlyLeuAlaGlyProProGlyGluSerGlyArgGlnGlySerProGlyAlaGlnGly--- 1011
 QY 411 CTTCATCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 355
 Db 1012 -----SerProGlyArgAspGlyAlaProGlyAlaLysGlyAspArgGlyGlnThr 1028
 QY 354 -----CCCCAGG-----GGCTG 340
 Db 1029 GlyProAlaGlyProProGlyAlaProGlyAlaProGlyAlaProGlyProValGly--- 1047
 QY 339 TAGAGGTGAGGACAAAGAGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 280
 Db 1048 -----ProAlaGlyLysAsnGlyAspArgGlyGlnThrGlyProAlaGlyProAla 1064
 QY 279 AGCTGCGGACCAAGCT----- 262
 Db 1065 GlyProIleGlyProAlaGlyAlaArgGlyProAlaGlyProGlnGlyProArgGlyAsp 1084
 QY 261 -----CTAACCGCA 253
 Db 1085 LysGlyGlnThrGlyGlnGlnGlyAspArgGlyLysGlyAlaArgGlyPheSerGly 1104
 QY 252 GCGCAGATCTCAGCAATCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 193
 Db 1105 LeuGlnGlyProProGlySerProGlySerProGlyGlnGlnGlyProSerGlyAlaSer 1124
 QY 192 GCTTCATGAGGCTCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 151
 Db 1125 GlyProAlaGlyProArgGlyProProGlySerAlaGlySerProGlyLysAspGlyLeu 1144

QY 150 -----CCTCTCCAGCCGCCCACTGCTCTCCAGAGG-----TCAGCA 112
 DB 1145 AsnGlyLeuProGlyProIleGlyProGlyProAlaGlyArgThrGlyAspSerGly 1164
 QY 111 GCTGCTGCGCCCTCATCTCTCTTTCGCCCGCCGCCCTGCGCCGCTGCTCT 55
 DB 1165 ProAlaGlyProProGlyProPro-----GlyProProGlyProProGlyPro 1180
 RESULT 8
 Q9VE45 PRELIMINARY; PRT; 950 AA.
 AC Q9VE45;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE CG7709 protein.
 GN CG7709.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blasej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter A.E., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Adiyani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bence P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borotova D., Botchan M.R., Bouck J., Brockstein P., Broctter P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downs M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkllov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Snie B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhao G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zheng F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL/ AE003723; AAF55584.1; -
 DR FlyBase: FBgn0038642; CG7709.
 DR InterPro: IPR002965; P rich extensn.
 DR PRINTS: PR01217; PRICHEXTEN.
 SQ SEQUENCE 950 AA; 91280 MW; B57AB0EA4D30D37 CRC64;

Alignment Scores:
 Pred. No.: 4,79e-18 Length: 950
 Score: 553.00 Matches: 282

Percent Similarity: 37.99% Conservative: 99
 Best Local Similarity: 28.12% Mismatches: 380
 Query Match: 9.68% Indels: 243
 DB: 5 Gaps: 43

US-10-032-996-6 (1-3121) x Q9VE45 (1-950)

QY 2854 ATGAGGCTCCAGTTCCTTTGG-----GCCATGGCCCAATTGTGATCCAGTGAGGG 2801
 DB 93 MetArgValGlnLeuPheGlyValIleAlaValAlaValAlaValSerLeu----- 110
 QY 2800 TACAGATGAGCGCGCTGATTTAGCGTGTGCGGTAGAGATCCCTTCACAGAGCAAAAC 2741
 DB 111 -----ValArgAspValIleGlyAlaGlnProPro-----ValAsnAsn 122
 QY 2740 AGCAGCTGCGCCCTTCAGCAGGCTGAGATCTCCCTCCAGAGGAGTCT----- 2690
 DB 123 AlATyrlleuProProSerSerProGlnArgProSerSerTyrlGlyAlaProProVal 142
 QY 2689 -----GCCCGAGGCTTACAGGGGTTTTCGACATCAATCTCCCGGCTG 2639
 DB 143 SerSerTyrlleuProProAla-SerGlyProAlaProSerPheAsnSerAlaProSerSe 162
 QY 2638 GCTGTGCGCAGTTCGCCCGCGGCTTGACAAAGCCCAAGTCCAGAGGAGCTGCC 2579
 DB 162 rSerTyrlAlaAlaPro-----SerGlnSerAlaSerSerGlyPr 176
 QY 2578 TCACCCACAGAGACCCGCCGCCAGGAGGCTCAGCCGCTCCCGCTGCC----- 2527
 DB 176 OTyPrAlaAlaAlaAlaProArgProSerSerTyrlGlyProAlaSerArgProSe 196
 QY 2526 -----TCAGTTCTCGGGGAGCCCT----- 2506
 DB 196 rSerSerTyrlGlyProProProSerArgProSerGlnSerTyrlGlyProProGlnAl 216
 QY 2505 -----GCTCTCCACAC-----ACAGCTGAGCGTTCCAGGTCAAGGTAAACGAGCG 2456
 DB 216 AlTyrlSerProIleHisArgAlaProSerSerSerTyrlGlyAlaProArgProAlaProBr 236
 QY 2455 TGACACAGTCCCGCGCTCTCTCCCGGAGTCCGG-----GGGCCCCCTCAGAGAGGGGAGAC 2399
 DB 236 cSerGlnSerTyrlGlyAlaProProSerAlaSerTyrlGlyProProlySerAlaProBr 256
 QY 2398 AGCAGAGAAAGCTCGGGTGCAGGTGAGTCACTTCATTTCCAGAGAGCTGTGCC 2339
 DB 256 cSerGlnSerTyrlGlyAlaProAlaProProSerSerTyrlGlyProProlySerAl 276
 QY 2338 TCCAGACACAGGCTCTCTCTCTCCACCCCTGCTGCGCTACAGGTTCAACACCGCG 2279
 DB 276 aProSerSerTyrlGlyAlaProAlaPro----- 286
 QY 2278 AGCTTCTGTTCAAGTCTGTTTATTCACATGTAACGGCCAGTTCCTTCAGAGCT 2219
 DB 287 -----AlaAlaProSer-----SerSerTyrlGlyAlaProAlaProProSerSe 302
 QY 2218 TGAATCCGGGCTCATCTACTGCTCTCTGCTGCTACCTTACACGAGTGTCTCA 2159
 DB 302 rTyrlGlyAlaProAlaAlaProSerSerTyrl----- 313
 QY 2158 CTCTGCTCCAGAGAGAGTGCATGTTCTGCTGCTGCTCTCTGCTGAGGTCAGGTG 2099
 DB 314 -GlyAlaProAlaAlaProSerSerSerTyrlGlyAlaProAlaAlaProSerSerTy 333
 QY 2098 CGGT-----CATTCAGAGCGCTCGCTCCAGGCGCACCTCCACCGAGTAC 2051
 DB 333 rGlyAlaProAlaProProSerTyrlSerTyrlGlyAlaProAlaProProSerSerTy 353
 QY 2050 ACCAGGCTCTGCTGCTCTCTCCAGTCTCCAGTTCAGAGAGCAATCTGCTG 1991
 DB 353 rGlyAlaProAlaAlaProSerSerTyrlGlyAlaProAlaProProSerSerTy 373
 QY 1990 TCTGCTCTCTCGAGGCTACCACTTGTCAAAATATTCTTCAGAGAGGCTGTGCT 1931

RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Flybase:
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF034486; AAF48140.3;
 SQ SEQUENCE 1099 AA; 94352 MW; 1A24A01379BA0AD1 CRC64;

 Alignment Scores:
 Pred. No.: 1,64e-17 Length: 1099
 Score: 541.50 Matches: 299
 Percent Similarity: 35.15% Conservative: 69
 Best Local Similarity: 28.56% Mismatches: 452
 Query Match: 5.48% Indels: 227
 DB: Gaps: 50

 US-10-032-996-6 (1-3121) x Q8IR84 (1-1099)
 QY 2832 CCATGCCCCCAATTGTGATGCC-----AGTAGGGTACAGATGAGGCTTG 2785
 DB 6 ProLeuProGlyAlaProGluProPheCysProGlySerArgGlyLeuGlnValGlyGly 25
 QY 2784 GATTAGGATGCGGTAGAGACTGCCCTTCACAGAAAGCAAGCAGCAGCTGCCCTTT 2725
 DB 26 ThrIleGlyMetPheGlyArg-----GlyCysGlyGly 36
 QY 2724 CAGCAGCTCGAGCTCCTCCACAGAGAGGCTGCCCGGAGGCTTAA----- 2674
 DB 37 GlnGlu-----ProAlaProProGlySerThrGlyIleAspGlyThrGlnProAla 54
 QY 2673 GGGGGTTTTCGGACATCATCATCCCGGGGCTGCTCGCTGCA-----GTTCC 2623
 DB 55 GlyIlePheGlyProArgThrAspAspProGlyIleuProAlaGlyThrLeuGlnPro 74
 QY 2622 GCGGGGCTTGACAAAGGCCCAAGTCCAGGGCAGGCTGCTCAACCA----- 2572
 DB 75 AlaGlyAlaProProAlaGlyThrLeuGlnSerIleGlyAlaProProAlaGlyThrLeu 94
 QY 2571 CAGAGACACCGCCACACAGGGGCTCAGCGGCTCGCTCAGTTCAGTTCCTCGGG 2512
 DB 95 GluSerAlaGlyAlaProProAlaGlyThrLeuGlnProAlaGlyAlaProProAlaGly 114
 QY 2511 ACCCTGCTCTCCACACACAGGCTCAGGCTTCCAGGTCACAGGTAACGAGCTGA 2452
 DB 115 Thr-----LeuGlnSerAlaGlyAlaProProAlaGlyThrLeuAlaSerAlaGlyAla 132
 QY 2451 CCAAGTCCCGGCTCTCCCGGCGGCGGCGGCGGCGGCTCAGTAGAGGGGAGCAGCAG 2392
 DB 133 ProProProGlyThrLeuGlnSerAlaGlyAlaProPro-----AlaGlyThrLeuAla 150
 QY 2391 GAGCTCGGGTCCAGGTGAGCTCATCTTCATTTCAGAGAGCTGTCTCCGAGC 2332
 DB 151 SerAlaGlyAlaPro-----ProAlaGly 158
 QY 2331 ACAAGCTCTCTTCCCGACCGCCGCTGCTGAGCTCAGGCTCAGACCGCGAGCTTCT 2272
 DB 159 ThrLeuGlnSerAlaGlyAlaProProAlaGlyThrLeuGlnSer-----AlaGly 175
 QY 2271 GTTTCAGTTCCTGTTTATTCACATGTAACGGCCAGTTCCTTCCAGACTTGAATCC 2212
 DB 176 AlaProProAlaGlyThrLeuGlnSerAlaGlyAlaProProAlaGlyThrLeuGlnSer 195
 QY 2211 ---GGGCTCATACGCTCCTGCTGCTGCTTACCCCTTACCGAGGTGCTGCTGCT 2155
 DB 196 AlaGlySerProProAlaGly-----ThrLeuGlnProAlaGly---AspPro 210
 QY 2154 GCTGAGAGCAGCTGATGTTCTGCTGCTGCTGCTGCTGAGGCTCAGCTGCGGT 2095
 DB 211 ProAlaGlyGlyAspPro---SerAlaGlyGlnProGlyAlaAlaGlnAspProSerAla 229

QY 2094 CCATCTCCAGGCGCTGCC-----GCTCCAGGGCCACCTCCAGCAGTAGACCA 2047
 DB 230 GlyAlaProGlyAlaAlaValAlaThrSerAlaGlyGlnProGlyAlaAlaGlnAspPro 249
 QY 2046 GCTCTCTGCTCTCTCCAGCTGCTCAGTTCGAGAGGCAATCTGCTGCTGCT 1987
 DB 250 SerAlaGlyAlaProGlyAlaAlaValAlaThrSerAlaGlyGlnProGlyAlaAlaGln 269
 QY 1986 GCTCTCTGAGGCTCAGCTCCTTCAAAATCTTGACAGAGGCGCTGCTGCTG 1927
 DB 270 AspProSerAlaGlyAlaProGlyAlaAlaAspAspProSerAlaGlyAla----- 286
 QY 1926 AGATGAGAGTAGTCTGAGCTTGCCCTGAGGTTCACTTCGCACTGGACAGCAAGAG 1867
 DB 287 -----ProGlyAlaAlaValAlaThrSerAlaGlyAlaPro 298
 QY 1866 CTGAGGCCCAAGACCCCGCTGGCGGCATGTATGGCCCTATTCTTAACTAATGGAG 1807
 DB 299 GlyAlaAlaGlnAspProSerAlaGly----- 307
 QY 1806 CATCCAGGCGCTCGATGAGCTCATCCACTGCA-----ACA 1771
 DB 308 GluProGlyAlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValAlaThrSer 327
 QY 1770 GCGTCCGCTCTCTCGGGGACAGACACTCCCTGCTCAGCTTGCCTGATCTCA 1721
 DB 328 AlaGlyAlaPro-----GlyAlaAlaAspAspProSerAlaGly-----AlaPro 342
 QY 1710 GCGCTGCTGAGCAGAGTCTTCTCTCGGCGAGGCTGTCATCTCCCGCGGATCT 1651
 DB 343 GlyAlaAlaValAlaThrSerAlaGlyGlnProGlyAlaAlaGlnAspProSerAlaGly 362
 QY 1650 GCTGCTGCTCTGCGGCGCTGCCCGCAGCTGCCCTCTTCCGACAGCTCTTCT 1591
 DB 363 AlaProGlyAlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValAlaThrSer 382
 QY 1590 CCAGGCTCAGCGGCTGACATCGACAGATGCTCCTGTTAGAGG---CCTGAGTGG 1534
 DB 383 AlaGlyAlaProGlyAla-----AlaGlnAspProSerAlaGlyGlnProGly--- 398
 QY 1533 ATCTCAGCGCTTCTCTCCAGCCCGCTTCTCTCTGATCATCAGGCTCTCTTGGCCA 1474
 DB 399 -----AlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValAlaThrSerAla 416
 QY 1473 GGATGCTCCCGCTTGTGAGTCTCTCCCGCAGCTCCTCCAGCGCGCGCTGCTGTA 1414
 DB 417 GlyIleProGlyAla-----AlaGlnAspProSerAlaGlyAlaProGlyAlaAlaVal 434
 QY 1413 -----GCACTTCTCATCTCTGCTCCAGCCTTCT 1381
 DB 435 ValThrSerAlaGlyGlnProGlyAlaAlaGlnAspProSerAlaGly---AlaProGly 453
 QY 1380 TCTGCTCTCATCTCTGCTGCTGCTTCCAGGCTGACACAGAGCGGTGCTGCTGCT 1321
 DB 454 AlaAlaAspAspProSerAlaGlyAlaProGly-----AlaAlaVal 467
 QY 1320 GCTCTTCTCTGAGAGCGCGCATCTCTCCGCTTAATCTTCAGAGATCTTCTGCT 1261
 DB 468 ValThrSerAlaGlyGlnProGlyAlaAlaGlnAspProSerAlaGlyAlaProGlyAla 487
 QY 1260 GCT---CATGTTTCAGCTCAGGCTCTTACCGGGGCTGCGC---GCTTGCATTTG 1207
 DB 488 AlaAspAspProSerAlaGlyAlaPro-----GlyAlaAlaValAlaThrSerAlaGly 505
 QY 1206 CCTCAGGCGCGGCTTGTGCTCGCTCTCCGAGAGCGCGCTGCTGCACTGCTGCT 1147
 DB 506 GluProGlyAlaAlaGlnAspProSerAlaGlyGlnProGlyAlaAlaAspAspProSer 525
 QY 1146 GCTGCCGAGAGCTGAGCT-----TCCGCTGAGCT 1114
 DB 526 AlaGlyAlaProGlyAlaAlaValAlaThrSerAlaGlyGlnProGlyAlaAlaGlnAsp 545
 QY 1113 CCTCAGTCTGCTCTCAGCTTGGGCGACAGTAGACACAGCGCT---CCGTAGCTGCT 1057


```

Db      546 ProSerAlaGlyAlaProGlyAlaAlaValThrSerAlaGlyGluProGlyAlaAla 565
      1056 TCTTCTCTTCAAGCAGCTGACCTGCTGGGCGGACGACCTCTCTGCGAATCTCT 997
      566 GluAspProSerAlaGlyAlaProGlyAlaAlaAspAspProSerAlaGlyAlaPro 584
      996 GAGGCGGAG-----ACCGCTCGCAGCATCTCGAGGCT-----964
      585 GlyAlaAlaValThrSerAlaGlyGluProGlyAlaAlaGluAspProSerAlaGly 604
      963 -----CCTTGCCCTCGAGCTCCCGGAGCTGCTGCGCTTCAAC 925
      605 GluProGlyAlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValThrSer 624
      924 TCAAGCTCGGCGGACCTGCTCTG-----CCTCTGCTCCAGCTCCGGATAGCGTGGC--- 871
      625 AlaGlyGluProGlyAlaAlaGluAspProSerAlaGlyAlaProGlyAlaAlaAspAsp 644
      870 ----TGCTGCTGCGGCTTCAAGGCTGAGCTGCTTCTGCTGCGGACGAGCTCGCCATTA 814
      645 ProSerAlaGlyAlaProGlyAlaAlaValThrSerAlaGlyGluProGlyAlaAla 664
      813 GCTCTCTCTTCAATGCGGATCTATGACCAAGCTCCCGGATCTTCTGCTGCGGCT----- 760
      665 GluAspProSerAlaGly-----AlaProGlyAlaAlaAspAspProSerAla 680
      759 ---GGGCGGACCGGCTGAGCTGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 703
      681 GlyAlaProGlyAlaAlaValThrSerAlaGlyAlaProGlyAlaAlaGlu----- 698
      702 CTTGCTCCCAACCACTGCTGAGACCTGGAATGCTGATCCACTCTCTCAAGCATA 643
      699 ---Asp---ProSerAlaGlyGluProGlyAlaAlaAspAspProSerAla----- 713
      642 GCTCTGCGGCTCTCTCTGCGAGACTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 583
      714 -----GlyAlaProGlyAlaAlaVal----- 721
      582 TGCTGATCTATTCTTGGCGAGGTGTAAGGTCCGCTGCGGAGGCTCTCTCTCTCT 523
      722 -----ThrSerAlaG 725
      522 CTCTCTGAGGAGCTGAGAGGCA-----CTTCCAGCCTGCTCACTGATCA 472
      725 GlyGluProGlyAlaAlaGluAspProSerAlaGlyAlaProGlyAlaAlaAspAspPro 745
      471 GCAACTCAGCTCCAGCTCCCTCCATTTGTCACCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 412
      745 exAlaGlyAlaProGlyAlaAlaValThrSerAlaGlyGluProGlyAlaAlaGlu 765
      411 CTTATCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 352
      765 spsSerSerAlaGlyGluProGlyAlaAlaAspAspProSerAlaGlyAlaProG 783
      351 CCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 304
      783 GlyAlaAlaValThrSerAlaGlyGluProGlyAlaAlaGluAspProSerAlaGly 802
      303 CATTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 253
      803 AlaProGlyAlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValThrSer 822
      252 GCCGAGGAGTCCA-----CATCT-----CCT 232
      823 AlaGlyGluProGlyAlaAlaGluAspProSerAlaGlyAlaProGlyAlaAlaAspAsp 842
      231 CTGCTGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 172
      843 ProSerAlaGlyAlaProGlyAlaAlaValThrSerAlaGlyAlaProGlyAlaAla 862
      171 CAGCGAAGAGTCTCGATTCTCT-----CCTCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 127

```

```

Db      863 GluAspProSerAlaGlyGluProGlyAlaAlaAspAspProSerAlaGlyAlaProGly 882
      126 TCTGAGGAGTCAAGCAGCTGCTGCGGCGGCT-----CATGCTCTTCTGCGGCGGAG 76
      883 AlaAlaValThrSerAlaGlyGluProGlyAlaAlaGluAspProSerAlaGlyAla 902
      75 CCGGCTGCGGCGGCTGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 16
      903 ProGlyAlaAlaAspAspProSerAlaGlyAlaProGlyAlaAlaValThrSerAl 922
      15 CCGAGCTCAGGAGC 3
      922 adGlyAlaProGly 926

RESULT 11
ID 076045 PRELIMINARY; PRT; 1461 AA.
AC 076045;
DT 01-NOV-1998 (TRIMBLrel. 08, Created)
DT 01-NOV-1999 (TRIMBLrel. 12, Last sequence update)
DE 01-OCT-2002 (TRIMBLrel. 22, Last annotation update)
DE Pro alpha 1(I) collagen.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=85130970; PubMed=2857713;
PA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, Alu repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=88329734; PubMed=2843432;
PA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:105-115(1988).
[3]
SEQUENCE FROM N.A.
RP MEDLINE=9025644; PubMed=3178743;
PA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
PA Jernisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
[4]
SEQUENCE FROM N.A.
RP MEDLINE=91138770; PubMed=1995349;
PA Maier A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
[5]
SEQUENCE FROM N.A.
RP MEDLINE=92157916; PubMed=1787829;
PA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
PA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
[6]
SEQUENCE FROM N.A.
RP MEDLINE=98107942; PubMed=9443882;
PA Korkko J., Ala-Korkko L., De Paese A., Nuytincx L., Barley J.,
PA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
RT identification of common sequences of null-allele mutations.";

```

RL Am. J. Hum. Genet. 62:98-110(1998).
 RN [7]
 RP SEQUENCE FROM N.A.
 RA Korkko J.M., Earley J.J., Nuytink L., DePaep A., Prockop D.J.,
 RA Ala-Kokko L., (May-1999) to the EMBL/GenBank/DBJ databases.
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF017178; AAB94054.2; -
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR000885; Fib_collagen_C.
 DR InterPro: IPR01007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; VWC; 1.
 DR Prodom; PD000007; Collagen; 2.
 DR Prodom; PD02078; Fib_collagen_C; 1.
 DR SMART; SM0038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C; 1.
 KW Collagen.
 SQ SEQUENCE 1461 AA; 138630 MW; 9ACF5DE30EA78E21 CRC64;

 Alignment Scores:
 Pred. No.: 1,84e-17 Length: 1461
 Score: 540.50 Matches: 290
 Percent Similarity: 30.49% Conservative: 57
 Best Local Similarity: 25.48% Mismatches: 424
 Query Match: 9.46% Indels: 367
 DB: 4 Gaps: 58

 US-10-032-996-6 (1-3121) x 076045 (1-1461)
 QY 2847 TCCAGTCTTTTGGCCATGCGCCCAATTGTTGATCCAGTGAAGGTACAGATGAGGC 2788
 DB 211 SerGlyPrometGlyProArgGlyProProGlyProProGlyLysAsnGlyAspAspGly 2310
 QY 2787 CTGATTAGGTGGTGGCGGTAAAGACTGCTTACACAGAACAAACAGGACGTGCCCC 2728
 DB 231 ---GluAlaGlyLysProGlyLysProGly---GluArgGlyProProGlyPro- 246
 QY 2727 TTTCAAGCAGCTCGAGAGTCTCCCTCCAGGAGGTGCCCCGAGGCTTACAGAGG- 2671
 DB 247 ---GlnGlyAlaArgGlyLeuProGlyLysAlaGlyLeuPro---GlyMetLysGlyHis 264
 QY 2670 ---GGTTTTCGGAATCATCATCCTCCCGGCTGCTGCTGCGAGTTCCCGCGGGGCT 2614
 DB 265 ArgGlyPheSerGlyLeuAspGlyAlaLysGly---AspAlaGlyProAlaGly--- 281
 QY 2613 TGGACAAAGGCCCAAGTTCAGGGGAGGCTGCTCACCACAGAAAGCAGCCGCCCA 2554
 DB 282 -----ProLysGlyLysProGlySer 288
 QY 2553 CCAAGGAGCTCAGCCGCTCCGCTGCTCAGTTCTCGGGGAGCCCT- 2506
 DB 289 ProGlyGluAsnGlyAlaProGlyGlnMetGlyProArgGlyLeuProGlyLysArgGly 308
 QY 2505 ---GCTCTCACCACACAGAGCTCAGCGTTCCAGGTCAAGGTAAAGGAGCGT 2455
 DB 309 ArgProGlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAlaAla 328
 QY 2454 GAGACA-----AGTCCCGCTCTCTCCCGGGTGGCGGGGCCCCCTCAGTGA 2407
 DB 329 GlyProProGlyProThrGlyProAlaGlyProProGlyPheProGlyAlaValGlyAla 348
 QY 2406 GGGGGAGCAGCC-----AGAGAGCTCGGCTGCTCCAGGTGAGCTCATCTT 2362
 DB 349 LysGlyGluAlaGlyProGlnGlyProArgGlySerGlnGlyProGlnGlyValArg--- 367
 QY 2361 CATTTCCAGAGCTGTCTGCTCCTCCAGACAGAGCTCTCTTCTCC-----CACCCC 2308
 DB 368 -----GlyGluProGlyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnPro 385
 QY 2307 TGGTGGCTACAGGCTTACACAGCCGAGGCTCT-----GTTTCAGTT 2263

DB 386 GlyAlaAspGlyGlnProGlyAlaLysGlyAlaAsnGlyAlaProGlyLysAlaGlyAla 405
 QY 2262 CCTGGTTTTCACATGTAAGGCCCA-----GTTCCCTCT 2227
 DB 406 ProGlyPheProGlyAlaArgGlyProSerGlyProGlnGlyProGlyGlyProProGly 425
 QY 2226 CCAGAGCTTGAATCCGGGCTCATCTGCTCTGCTGCTGCTGCTGCTTACACCTTACCCGAGGT 2167
 DB 426 ProLysGlyAsnSerGlyGluProGlyAlaPro- 436
 QY 2166 GGTCTGACTGCTGCGACAGACAGCTGATCTTCTGCTGCTTCTGCGCAGGG 2107
 DB 437 GlySerLysGlyAspThrGlyAlaLysGly-----GluProGlyProValGly 452
 QY 2106 TCAGCTGGCGGCTCATCTCCAGGC----- 2083
 DB 453 ValGlnGlyProProGlyProAlaGlyGlnGlnGlyLysArgGlyAlaArgGlyGluPro 472
 QY 2082 -----GTCGCCGCTCCAGGCGCAGCT---CCAGCAGTACACAGCCTGCTGCTCT 2032
 DB 473 GlyProThrGlyLeuProGlyProProGlyGluArgGlyGlyProGlySerArgGlyPhe 492
 QY 2031 CCAAGTGA-----TCTCCAGTTCCAGAGGCAATCTGCTCT 1993
 DB 493 ProGlyAlaAspGlyAlaAlaGlyProLysGlyProAlaGlyGluArgGlySerProGly 512
 QY 1992 GGTGCTGCTCTCTCGAGGCTCAGCCTTGTCAAAATACCTGACAGAGAGGCTGCTGT 1933
 DB 513 ProAlaGlyProLysGlySerProGlyGluAlaGlyLysArgProGlyGluAlaGlyLeuPro 532
 QY 1932 TCTGTGAG---ATGAGAGTACGTAGCTTGCGCATGAGTTCATCTGCACTGAGACA 1876
 DB 533 GlyAlaLysGlyLeuThrGlySerProGlySerProGlyProAlaGlyLysThrGly--- 551
 QY 1875 GCAACGAGGCTGAGCGCCGAGACCCGCTGCGCGCATGTGATGCTTACTTACT 1816
 DB 552 -----ProProGlyProAlaGly- 557
 QY 1815 CATTCGACATCCAGGAGGCTCGATGCGCTCATCTCAATGAAACAGAGTCCGCTCT- 1759
 DB 558 GlnAspGlyArgProGlyPro-----ProGlyProProGlyAlaArgGlyGlnAlaGly 575
 QY 1758 -----CCTCGGGAGACAGCACTCCCTGCTCCTGAGCTGCGTCA 1717
 DB 576 ValMetGlyPheProGlyProLysGlyAlaAlaGlyGluProGlyLysAlaGlyLysArg 595
 QY 1716 TCT-----CCAGGCTGCTGAGCAGCAGTCTCTCTGCGCAGGCTGT 1669
 DB 596 GlyAlaProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGluAlaGlyAla 615
 QY 1668 CGATCTCCCGCGATCTGCTGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCT 1609
 DB 616 GlnGlyPro- 618
 QY 1608 TCTCGACAGCTCTTCTTCAGAGTCTCCAGCGGCTGACACATCGACGATCTCTGT 1549
 DB 619 -----ProGlyProAlaGlyProAlaGlyGluArgGlyGlnGlnGlyProAla 634
 QY 1548 TGAGGCTGCTGATCTCAGGCTGCTGCTGCTTCCAGCCCGCTTCTCTGATCAAGG 1489
 DB 635 GlySerProGlyPhe---GlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGly 653
 QY 1488 CCTCTTCTTGGCAGAGTGCCTCCGCTGCTGAGAGCTCTCTCCCGCAGCT----- 1438
 DB 654 LysProGlyGlnGlnGlyValProGlyAspLeuGlyAlaProGlyProSerGlyAlaArg 673
 QY 1437 -----CCTCCAGGCGCCGCGCTCT 1417
 DB 674 GlyGluArgGlyPheProGlyGluArgGlyValGlnGlyProProGlyProAlaGlyPro 693
 QY 1416 GTAGCAGCTTCTTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1357
 DB 694 ArgGlyAlaAsnGlyAlaProGly-----AsnAspGlyAlaLysGlyAspAlaGly 710

QY 1356 GTTCCAGGCTGACCAAGACCCTTGTCTGCACTGCGCCTTCTCTTGGAGGCG--- 1300
 Db 711 AlaProglyAlaPro-----GlySerGlnGlyAlaProgly 722
 QY 1299 ---CGATCTCTTCCGCTTATCTTGAAGATCTTCTGCTGCTTCACGCTTCACTCA 1243
 Db 723 LeuGlnGlyMetPro-----GlyGluArgGlyAlaAlaGlyLeuProglyPro 738
 QY 1242 GCTCTGACGCGGTGCTGCGGCTGCTCA----- 1213
 Db 739 LysGlyAspArgGlyAspAlaGlyProLysGlyAlaAspGlySerProglyLysAspGly 758
 QY 1212 -----TTTCTGCTTCCAGGCGCGCTTCTGCTCTCTCTCTGCG 1174
 Db 759 ValArgGlyLeuThrGlyProIleGlyProIleGlyProAlaGlyAlaProglyAspLys 778
 QY 1173 GAAGCGGCTGTGAGCTGTCCCTGCTGCTGCGCATGACGCTCCGCTCGAGCT 1114
 Db 779 GlyGluSerGlyProSerGlyProAla----- 787
 QY 1113 CTGCGAGCTGCTTCTCACTGTGCGCGACAGTACACGAGCGCTCGTAAGCTGCTCT 1054
 Db 788 -----GlyProThrGlyAlaArgGlyAlaProglyAspArgGly 800
 QY 1053 TCTGCTTACAGACTGCA-----CTGCGCTGCGCGCGAGCGA 1015
 Db 801 GluProglyProProglyProAlaGlyPheAlaGlyProProglyAlaAspGlyGlnPro 820
 QY 1014 CCTCTCTGCGAACTCTCTGAGCGGAGACCGCTGCGACAGCATCTGAGACTCTTGGCT 955
 Db 821 GlyAlaLysGlyGluProglyAspAlaGlyAlaLysGlyAspAlaGlyProProglyPro 840
 QY 954 CGAGCTCCCGCA-----GCTGCTCTGCGCTTCACTCA 922
 Db 841 AlaGlyProAlaGlyProProglyProIleGlyAsnValGlyAlaProglyAlaLysGly 860
 QY 921 GCTCGCGCGGCACTGTCTGCTCTCTGCTCAAGCTCCCGATACCTGCTGTGCTGCG 862
 Db 861 AlaArg---GlySerAlaGlyProProgly---AlaThrGlyPheProglyAlaAlaGly 878
 QY 861 GCTTCAGGCGCTGAGCTGCTTCTCTGCGGAGACGAGTCCCAATAGCTCTCTCTCA 802
 Db 879 ArgValGlyProPro-----GlyProSerGlyAsnAlaGlyProProgly 893
 QY 801 TGGCGATGTGATAGCCAGCTCCC---GGAATCTTCTGCTGCGCTGCGCGAGCGGCACT 745
 Db 884 Pro-----ProglyProAlaGlyLysGlnGlyLysGlyProArgGly--- 908
 QY 744 CTGAGGCTGTGCGAGGGGAGCTGCGGG---CCTGAATCTGCGGCTTCTCTCCAC 691
 Db 909 -----GluThrGlyProAlaGlyArgProglyGluValGlyPro----- 921
 QY 690 CAACGCTGTGAGACCTGGAAAGGTGATCCAACTCTCAAGGCAAGCT----- 640
 Db 922 -----ProglyProProglyProAlaGlyGluLys 931
 QY 639 -----CTGCGCCTTCTCTCTGCGACACTCCCTGCGCGTGGCGCCCTCTGACTGC 586
 Db 932 GlySerProglyAlaAspGlyProAlaGlyAlaProglyThrProglyProGlnGly--- 950
 QY 586 AGTTCGATCTCTATTTCTGCGCGAGGTAGAGTCCGCGCGGCTGCTCTCT--- 532
 Db 951 -----IleAlaGlyGlnArgGlyValAlaGlyLeuProglyGlnArg 964
 QY 531 -----CCTCTCTTCTCT-----CCTTGAAGCAGCTGAG 502
 Db 965 GlyGluArgGlyPheProglyLeuProglyProSerGlyGluProglyLysGln----- 982
 QY 501 AGCCACTTCCGAGCTGTTCACCTGACGAGCAACT---CAGCTCCAGCCTCCCTGCGCAT 445
 Db 983 -----GlyProSerGlyAlaSerGlyAlaArgGlyProProglyPro 956

QY 444 TTGTCACT-----CCTCTCCCTCTGTGTCAGGCAACTTCACTCCAGGAGGC 394
 Db 997 MetGlyProProglyLeuAlaGlyProProglyGluSer-----Gly 1010
 QY 393 AGGAGCGCGCACCA----- 379
 Db 1011 ArgGlnGlyAlaProglyAlaGlnGlySerProglyArgAspGlySerProglyAlaLys 1030
 QY 378 -----TGCCAGCACATGGGCTGGGCACTCC-----CGAGGGGG 343
 Db 1031 GlyAspArgGlyGluThrGlyProAlaGlyProProglyAlaProglyAlaProglyAla 1050
 QY 342 CTGTATGAGATC---GAGGCAAAAGAACCCGAGAGGAGGCACTTCAGAGCCGAGGC 286
 Db 1051 ProglyProValGlyProAlaGlyLysSerGlyAspArgGlyGluThrGlyProAlaGly 1070
 QY 285 CCCCCAGCGCTGGCGCACCA----- 265
 Db 1071 ProAlaGlyProValGlyProValGlyAlaArgGlyProAlaGlyProGlnGlyProArg 1090
 QY 264 -----GCTCTA 259
 Db 1091 GlyAspLysGlyGluThrGlyGlnGlnGlyAspArgGlyIleLysGlyHisArgGlyPhe 1110
 QY 258 ACCGAGCGGCACTTCCACCATCTCTCTGCTGCTCAAGCAGCGGCTGCTGCAATT 199
 Db 1111 SerGlyLeuGlnGlyProProglyProProglyLysSerProglyGlnGlnGlyProSerGly 1130
 QY 198 TGTACTGCTTCAGAGGGGCTCTCCAGGCGGAGGCAAAAGTCTGCTCT----- 151
 Db 1131 AlaSerGlyProAlaGlyProArgGlyProProglySerAlaGlyAlaProglyLysAsp 1150
 QY 150 -----CCTCTCAGCGCGGCACTGTTCTGACAGG---TCAGCACT 109
 Db 1151 GlyLeuAsnGlyLeuProglyProIleGlyProProglyProArgGlyArgThrGlyAsp 1170
 QY 108 GCTCGCGCGCTCATCTCTTTCGCGCGCGCGCGCGCGCGCGCGCTGCTCT 55
 Db 1171 AlaGlyProValGlyProProglyProProglyProProglyProProglyPro 1188

RESULT 12
 OSN473 PRELIMINARY; PRT; 1464 AA.
 ID OSN473
 AC 08N473
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN NCBI_taxonomy=9606;
 RX (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL, BC036531, AAH36531.1; -
 DR InterPro; IPR000087, Collagen.
 DR InterPro; IPR000885, Fib_collagen_C.
 DR InterPro; IPR001007, VWF_C.
 DR Pfam; PF01410, COLF1; 1.
 DR Pfam; PF01391, Collagen; 18.
 DR ProDom; PD000007, Collagen; 2.
 DR ProDom; PD0002078, Fib_collagen_C; 1.
 DR SMART; SM00038, COLF1; 1.
 DR SMART; SM00214, VWF; 1.
 DR PROSITE; PS01208, VWF; 1.
 KW Hypothetical protein; Collagen.
 SQ SEQUENCE 1464 AA; 139011 MW; B0581F8D1C89DDE8 CRC64;

Alignment Scores: 1.84e-17 Length: 1464
 Pred. No.:

Score: 540.50 Matches: 290
 Percent Similarity: 30.49% Conservative: 57
 Best Local Similarity: 25.48% Mismatches: 424
 Query Match: 9.46% Indels: 367
 DB: 4 Gaps: 58

US-10-032-996-6 (1-3121) x Q8N473 (1-1464)

2847 TCACGTTCTTTGGGCGCATGGCCCAAAATTGTTGATCCCAAGTAGAGTGAAGGCG 2788
 214 SerGlyProMetGlyProArgGlyProProGlyProProGlyProGlyAsnGlyAspAspGly 233
 2787 CTGGATTAGGGGTGGGTAGGAGCTGCTTACAGAGAAAGAAACAGGACAGCTGCCCC 2728
 234 ---GluAlaGlyLysProGlyLysProGly---GluArgGlyProProGlyPro--- 249
 2727 TTTCAGCAGGCTCGGAGTCTCCCTCCAAAGCAGGAGTCTGCCCGAGGCTTACAGG--- 2671
 250 ---GlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeuPro---GlyMetLysGlyHis 267
 2670 ---GGTTTTCGGGATCATCATCCCGGGGCTGCTCCGAGTTCCCGGGGCT 2614
 268 ArgGlyPheSerGlyLeuAspGlyAlaLysGly---AspAlaGlyProAlaGly--- 284
 2613 TGAGCAAGGCCCAAGTTCCAGGAGGCTGCTGCTACCCACAGAGAGACCCGCCCA 2554
 285 ---ProLysGlyLysProGlySer 291
 2553 CAGGGGCTCAGGCCCTCCCGCTCCATGTTCTCGGGGAGCCCT--- 2506
 292 ProGlyGluAsnGlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAlaAla 311
 2505 ---GCTCTCACCACACAGCTCAGCGTTTCCAGGTCCAGGTACAGGTAACGAGCGT 2455
 312 ArgProGlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAlaAla 331
 2454 GGACCA---AGTCCGCGCTCTCTCCCGGGTGGGGGGGCCCCCTCACTGA 2407
 332 GlyProProGlyProThrGlyProAlaGlyProAlaGlyPheProGlyAlaAlaGlyAla 351
 2406 GGGGGGACAGCC---AGAGAGCTCGGTCGCGCAGGTGAGTGAAGTCACTT 2362
 352 LysGlyGluAlaGlyProGlnGlyProArgGlySerGlnGlyProGlnGlyValArg--- 370
 2361 CATTTCCAGAGCTCTGCTCCCTCCAGACAGGCTCTCTCTCC---CACCCC 2308
 371 ---GlyGluProGlyProProGlyProAlaGlyAlaAlaGlyProAlaGlyAsnPro 388
 2307 TGCTGGGCTCAGCGTTCAACGCGGAGTCTCT---GTTTCAGTT 2263
 389 GlyAlaAspGlyGlnProGlyAlaLysGlyAlaAsnGlyAlaProGlyAlaAlaGlyAla 408
 2262 CCGTGTATTCACATGTAAGGCCCA---GTTCTCTT 2227
 409 ProGlyPheProGlyAlaAlaArgGlyProSerGlyProGlnGlyProGlyGlyProProGly 428
 2226 CAGAGCTGAATCCGCGCTCACTACTGCTCTGCTGCTGCTAACCCTTACCGAGGT 2167
 429 ProLysGlyAsnSerGlyGlnProGlyAlaPro--- 439
 2166 GGTCTGCACTCTCTCAGAGAGAGCTGATGTTCTGCTCGTCTCTTCTGCTGAGG 2107
 440 GlySerLysGlyAspThrGlyAlaLysGly---GluProGlyProValGly 455
 2106 TCAGCTGGCGCTCATCTCCAGG--- 2083
 456 ValGlnGlyProProGlyProAlaGlyGlnGlnGlyLysArgGlyAlaArgGlyLysPro 475
 2082 ---GCTCCGCTCCAGGCACT---CAAGCAATACACAGCTCTGCTCTCT 2032
 476 GlyProThrGlyLeuProGlyProGlyGlnArgGlyGlyLysSerGlyPhe 495
 2031 CAGCTGCA---TCTCAGTTCGAGAGAGCAATCTGCTCT 1993

496 ProGlyAlaAspGlyValAlaGlyProLysGlyProAlaGlyGluArgGlySerProGly 515
 1992 GGTGCTGCTCTCTCCGAGCGTCACCACTTGTCAAAATCTTGAGAGAGGAGGCTCGG 1933
 516 ProAlaGlyProLysGlySerProGlyGlnAlaGlyArgProGlyGlnAlaGlyLeuPro 535
 1932 TCTCTGAG---ATGAGAGTACTGAGCTTGCCCATGAGATTCATCTCGCATGGACA 1876
 536 GlyAlaLysGlyLeuThrGlySerProGlySerProGlyProAspGlyLysThrGly--- 554
 1875 GCAACGAGGCTGAGGCCGAGACCCGCTGGCGGCAATGATGCTCATCTTACT 1816
 555 ---ProProGlyProAlaGly--- 560
 1815 CAATGCAAGCATCAGGAGCTCATGCTCATCACTCAAGACAGAGCTCCGCTCT--- 1759
 561 GlnAspGlyArgProGlyPro---ProGlyProProGlyAlaAlaArgGlyGlnAlaGly 578
 1758 ---CCTCGGGAGACAGCACTCCCTGCTCCTCAGCTCCGCTGCA 1717
 579 ValMetGlyPheProGlyProLysGlyAlaAlaGlyGluProGlyLysAlaGlyLysArg 598
 1716 TCT---CAGGCGCTGCTGAGCAGCAGTCTTCTCTGCGCAGAGCTGT 1669
 599 GlyValProGlyProProGlyAlaAlaGlyProAlaGlyLysAspGlyGlnAlaGlyAla 618
 1668 CGATCTCCGCGAGTCTGCTGCTGAGCTGAGGAGCTGCTGCGCAGCTCCGCTCT 1609
 619 GlnGlyPro--- 621
 1608 TCTGAGACAGCTCTTCTCCAGGTCCTCCAGCGGCTGAGCACTCGACATGCTCTGT 1549
 622 ---ProGlyProAlaGlyProAlaGlyGluArgGlyGlnGlnGlyProAla 637
 1548 TGAGGCTGCTGATCTCAGGCTGCTGCTCCAGCCCGCTTCTCTCTGATCAGG 1489
 638 GlySerProGlyPhe---GlnGlyLeuProGlyProAlaGlyProProGlyGlnAlaGly 656
 1488 CTTCTTCTGAGAGATGCTCCCGCTTGTGAGAGCTCTCCCGCAGCT--- 1438
 657 LysProGlyGlnGlnGlyValProGlyAspLeuGlyAlaProGlyProSerGlyAlaArg 676
 1437 ---CCTCCAGCGCCCGCTGCT 1417
 677 GlyLysArgGlyPheProGlyGluArgGlyValGlnGlyProProGlyProAlaGlyPro 696
 1416 GTAGACCTTCTCATCTCTGCTGTCAGACCACTTCTGCTCTCAATCTGCTGCT 1357
 697 ArgGlyAlaAsnGlyAlaProGly---AsnAspGlyAlaLysGlyAspAlaGly 713
 1356 GTTCAGGCTGACACAGAGCGTTGCTGTCACATGCGCTTCTCTGAGAGCGG--- 1300
 714 AlaProGlyAlaPro---GlySerGlnGlyAlaProGly 725
 1299 ---CATCTCTTCTGCTTATCTTCAAGATCTTCTGCTGCTATGCTTACGCTCA 1243
 726 LeuGlnGlyMetPro---GlyGluArgGlyAlaAlaGlyLeuProGlyPro 741
 1242 GCTCTTACGCGGCTGCGCTGCTGCTCA--- 1213
 742 LysGlyAspArgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGly 761
 1212 ---TTCGCTCCAGAGCGCGCTTCTGCTGCTGCTCCGCTCCGCT 1174
 762 ValArgGlyLeuThrGlyProLysGlyProProGlyProAlaGlyAlaProGlyLysPro 781
 1173 GAAGCGCTCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1114
 782 GlyLysSerGlyProSerGlyProAla--- 790
 1113 CTGCACTGCTTCTCACTCTGCGCCGACATGACACAGCGCTCTGCTGCTGCTCT 1054

2593 CAGGGCAGGCGCTGCTCCACCCACAGAAAGACACCCGGCCACACAGGGGGCTCAACGCGCTCC 253
Db ASPSerGlyProAlaGlyProIysGlyGlu-ProIysSerProGlyGluAsnGlyAlaPr 284
2533 CGGTGCGCTCAGTTCTCGGGGACCCCTGCTCTCACACACAGGCTCGAGCGTTTCAG 2474
284 oGlyGlnValAlaGlyProAlaGlyLeuSerGly----- 294
2473 GTCGAAGGTAAAGGACCGTGA---CCAATGCCGCGCTCTCT----- 2434
295 -----GluArgGlyAAspProGlyProSerGlyProAlaGlyAlaArgGlyAs 310
2433 -----CCCGGGTGGGGGGGGCCCCCTCACTGAAGGGGGACGACCGAAGACTCG 2384
310 nAspGlyAlaProGlyAlaAlaGlyProProGlySerThrGlyProSerGlyProProGly 330
2383 GGTCCAGAGTGGAGCTCAATCTTCATTTCCAGAGACCTGTGGCCCTCCG----- 2335
Db yPheProGlyGly-----ValGlyProIysGlyAspAlaGly 342
2334 -----AGCAGAGCTCTCTTCTCCACCACCCCTGCTGGGCTCAAGCGTTACAA 2285
342 yProGlnIysSerAAspGlyPro-----GlnGly 354
2284 CCGCCGAGCTTCTGTTTCAAGTTCTCTGGTTTATCCACATGAAGGCCCAAGTTCTTCC 2235
354 yGlyAAspGlyGluProGlyAlaProGlyGlnAlaGlyAlaAlaGlyProSerGlyAAspPr 374
2224 AGAGCTTGAATCCGGGCGCTCACTACGCTCTGCTGTCTACTAACCCTTCACCGAGAGTGG 2165
374 O-----Gly 375
2164 TCTCGACTCTGCTGACGAGACGCTGCATGTTCTGCTGCTCTCTTCTGCTGCAGGCTC 2105
375 ySerAAspGlyGluProGlyAlaAlaGlyAlaThrGly---AlaProGlyAlaAlaGlyAla 394
2104 AGCGGGGCGTCCA-----TCTCCAGAGGCGTGGCGGCTCCAGGGCACCTGCAGC 2057
394 aProGlyPheProGlyAlaArgGlyAlaProGlyAlaGlnGlyProGlyGlySerProGly 414
2056 CAGTACA----- 2050
414 yProIysGlyAAsnAsnGlyGluProGlyAlaGlnGlyAAsnIysGlyGluAlaGlyAla 434
2049 -----CAGAGCTCTGCT---GCTCTCCAGAGCGCATCTCCAGTTCCGAGAA 2006
434 sGlyGluProGlyProAlaGlyAlaGlnGlyProProGlyProSerGlyGluGlnGlyIly 454
2005 GCAATCTGCTGCTGTGCTGCTCTCTCCGAGCGCTCAACCTTGTCCAAATACTTGCAG 1946
454 sArgIysSerAAspGlyGluProProAlaGly-----ProProGlyProAlaGlyGluArg 472
1945 AGAGGGGCTCTGCTCTCTAGAGATAGAGCTAAGCTGAGCTTGGCGATGAGTTCACTCTCG 1886
472 gGlyGlyProGlySerAAspGlyPheProGlySerAAspGly-----AlaSerGlyPr 489
1885 CACTGGGACAGCAACAGAGCTGAGCCGGAAGCAACCCGCTGGCGCATGTGATGAGCTCA 1826
489 oIysGlyAlaProGlyGluArgGlyProAlaGlyProAlaGlyPro-----LysGly 506
1825 TTCTTAATCAATGGCAGACATCCAGG---CTTCGATGGCTTCATCCAACTGGAAACAC 1766
506 ySerSerGlyGlnIysGlyAAspGlyGluProGlyLeuProGlyAlaAlaGlyLeuThr 526
1768 GTCCGCTCTCTCTCCGGGGGACAGCAAC-----TCCCTGCGCTCAAGTTGCCGTG 1718
526 rGlySerProGlySerProGlySerAAspGlyLysThrGlyProAlaGlyAla 543
1717 ATCTCCAGGGCGCTTGAAGCAGCAAGTCT-----TCTCTTGGCGCA----- 1675
544 ---ProGlyGlnAAspGlyAArgAlaGlyProProGlyProProGlyAlaArgGlyGlnSe 562

[illegible]

[illegible]

RT	"Analysis of the mouse transcriptome based on functional annotation of RT 60,770 full-length cDNAs"; PL Nature 420:563-573 (2002). DR EMBL: AK041115; BAC30826.1; "-" SQ SEQUENCE 1464 AA; 136933 MW; 91F3246D90818449 CRC64;			
Alignment Scores:				
Pred. No.:	4,31e-17	Length:	1464	
Score:	532.50	Matches:	303	
Percent Similarity:	29.30%	Conservative:	48	
Best local Similarity:	25.29%	Mismatches:	452	
Query Match:	9,328	Indels:	35	
DB:	11	Gaps:	58	
US-10-032-996-6 (1-3121) x QEBLW4 (1-1464)				
QY	2709 CTCCTCCAGGACGAGGTCTGCCCCGAGGAGGCTTACAGG-----GGTTTTCGGACAT	2656		
DB	117 TLeProGlyAArgAsnGlyAAspPro---GlyLeuProGlyGlnProGlyLeuProGlyPro	135		
QY	2655 CAATCATCCCCGGGTGGCTGCTGGCAGTTCCCGCGGGGCTTGACA-----	2608		
DB	136 ProGlySerProGlyIleCysGluSerCysProThrGlyGlyGlnAsnTyrSerProGln	155		
QY	2607 -----AAGCCCAAGTTCCAGAGGAGGCTGCTTCACCCACAG	2569		
DB	156 PheAspSerTyrAspValIleSerGlyValIleGlyMetGlyGlyIleProGlyProAla	175		
QY	2568 GAAGCACCCGCCCCACGAGGGGCTCAGCCCTCCGCGGCTCACTTCTCGGGGAGC	2509		
DB	176 GlyProGlyProProGly-----ProProGlySerSerGlyHisProGlySer	192		
QY	2508 CCT---GCTCTCACACACACAGGCTCGAGCGTTCCAGAGTCAAGGATTA	2452		
DB	193 ProGlySerProGlyTyrGlnGlyProPro-----GlyGluProGlyGlnAlaGly	209		
QY	2451 CCAAGT---CCGCGCTCTCTCCGGGTGGGGGGGCCCTCAGTACGAGGGGAGCACCC	2395		
DB	210 ProAlaGlyProProGlyProProGlyValAlaGluGlyProAlaGlyProAlaGlyAsp	229		
QY	2394 AGAAGAGCTCGGGTCCACAGTGAAGTCACTTCATTCCAGAGAGGCTGTGCCCCCG	2335		
DB	230 GlyGluSerGlyAArgProGlyAArgPro-----GlyGluArgGlyLeuProGly	245		
QY	2334 AGCAGAGGCTCCTCTCTCCACCCCTGCTGGCTA-----CAGCTTCACACCGC	2281		
DB	246 ProProGlyIleTyrGlyProAlaGlyMetProGlyPheProGlyMetCysGlyHisArg	265		
QY	2280 -----CGAGCTTCTGTTTCAAGTTCTCTGTTTATCCACATGT	2245		
DB	266 GlyPheAspGlyAArgAsnGlyGlyIleGlyGlyGlyIleAlaProGlyLeuIleGlyIle	285		
QY	2244 AAGGCG-----CAAGTTCCTTCCGAGAGCTGAATCCGGGGCTCATCT	2200		
DB	286 AsnGlyLeuProGlyAAspAsnGlyAlaPro---Pro-----MetCysProArgGly	302		
QY	2199 GCTCTCTGCTGCTCTCAACCTTACCCAGAGT-----GGTCTGCAGCTCTGCTCAGGA	2146		
DB	303 AlaProGly-----GluArgGlyAArgProGlyLeuProGlyValAlaGly	317		
QY	2145 GCAGCTGCATGTTCTGCTCGAGTCTCTTCTGCTGACGAGGTCAGCTGGCGCTCATCTCA	2086		
DB	318 AlaArgGly-----AsnAspGlyAlaAlaArgIleSerGlyGlyAlnProGlyPro	333		
QY	2085 GGGCTGCGCTCCAGGCGCACT-----	2062		
DB	334 -----ProGlyProProGlyThrAlaGlyPheProGlySerProGlyAlaIle	349		
QY	2061 -----CCAGCCAGTACACACAGCGCTCTGCTGCTCTCTCAGCTCATCTCCAGTT	2014		
DB	350 GlyGlyValAlaGlyProAlaGlySerProGlySerAsnGlySerProGlyAlnArgGlyIle	369		
QY	2013 CGAGAGGCAATCTGCTCTGCTGCT-----GCTCTCTCGAGAGCTACACACTTGTACA	1957		

[illegible]


```

Db      1125 HisGlnGlyAlaIleGlySerProGlyProAlaGlyProArgGlyProValGlyProHis 1144
QY      183 CGTCCTCCAGCGCAGAAAGTCTCGGTTCTCTCTCCAGCCGCGCAGCTGTTCT 124
Db      1145 GlyProProGlyLysAspGlyThrSerGlyHisProGlyProIleGlyProProGlyPro 1164
QY      123 GCAGGTCA-----GCAGCTGCTGCGCCCGCCCTCATCT-----CCTTTC 85
Db      1165 ArgGlyAsnArgGlyLysArgGlySerGlyLysProGlyHisProGlyGlnProGly 1184
QY      84 GCCCGCGCGCCCTGCGCGCTGCTCTCGAC---GAGGCGCTCTGATGCCGCTAT 28
Db      1185 ProProGlyProProGlyAlaProGlyPro-CysCysGlyGlyGlyAlaAlaAlaIleAl 1204
QY      27 CCGGCGCGGAGCGGAGCTCAGGAGC 3
Db      1204 aglyvalglyglylulyssergly 1212

```

Search completed: January 29, 2004, 23:53:31
 Job time : 235 secs